THEORETICAL STUDIES ON PROTEINS TO REVEAL THE MECHANISM OF THEIR FOLDING AND BIOLOGICAL FUNCTIONS

A Dissertation

by

QIANG SHAO

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2009

Major Subject: Chemistry

THEORETICAL STUDIES ON PROTEINS TO REVEAL THE MECHANISM OF THEIR FOLDING AND BIOLOGICAL FUNCTIONS

A Dissertation

by

QIANG SHAO

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Yi Qin Gao
Committee Members,	Robert R. Lucchese
	Simon W. North
	Pingwei Li
Head of Department,	David H. Russell

December 2009

Major Subject: Chemistry

ABSTRACT

Theoretical Studies on Proteins to Reveal the Mechanism of Their Folding and Biological Functions. (December 2009)

Qiang Shao, B.S., Nanjing University;

M.A., College of William & Mary

Chair of Advisory Committee: Dr. Yi Qin Gao

The folding mechanism of several β-structures (e.g., β-hairpins and β-sheets) was studied using newly developed enhanced sampling methods along with MD simulations in all implicit solvent environments. The influence of different implicit solvent models on the folding simulation of β-structure was also tested. Through the analysis of the free energy landscape as the function of several suitable reaction coordinates, we observed that the folding of β-hairpins is actually a two-state transition. In addition, the folding free energy landscapes for those related hairpins indicate the apparent sequence dependence, which demonstrates different folding mechanisms of similar β-structures of varied sequence. We also found that the stability of backbone hydrogen bonds is determined by the turn sequence and the composition of hydrophobic core cluster in β-structures. Neither of these findings was reported before.

The processive movement of kinesin was also studied at the mesoscopic level. We developed a simple physical model to understand the asymmetric hand-over-hand mechanism of the kinesin walking on the microtubule. The hand-over-hand motion of

the conventional kinesin is reproduced in our model and good agreement is achieved between calculated and experimental results. The experimentally observed limping of the truncated kinesin is also perfectly described by our model.

The global conformational change of kinesin heads (e.g., the power stroke of neck-linkers which works as lever-arms during the kinesin walking, the transition between open and closed states of the switch region of the nucleotide binding domain in each head induced by the nucleotide binding and release) was studied for both dimeric and monomeric kinesins using a coarse-grained model, anisotropic network model (ANM). At the same time Langevin mode analysis was used to study the solvent influence on the motions of the kinesin head mimicked by ANM. Additionally, the correlation between the neck-linker and the nucleotide binding site was also studied for dimeric and monomeric kinesins. The former shows the apparent correlation between two subdomains whereas the latter does not, which may explain the experimental observation that only the dimeric kinesin is capable of walking processively on the microtubule.

ACKNOWLEDGEMENTS

First of all, I would like to express my gratitude to my advisor, Prof. Yi Qin Gao, for his guidance in my graduate study. It is Prof. Yi Qin Gao who led me to a new field of computational chemistry. As a great mentor, he taught me not only the scientific knowledge but also the positive attitude. Without him the work enclosed within this dissertation would not have been possible and without him I could not find out the right direction to follow in my future.

I am very appreciative of my committee members, Prof. Robert R. Lucchese, Prof. Simon W. North, Prof. Pingwei Li, for their time and effort to serve in my committee. Their advice was very valuable and appreciated.

Thanks also go to my friends and colleagues, Dr. Lijiang Yang and Dr. Yubo Fan. Both of them mentored me in some way or the other during my development as a graduate student: Dr. Lijiang Yang taught me how to perform molecular dynamics (MD) simulations and Dr. Yubo Fan shared his knowledge of quantum mechanics (QM) with me. I would like to thank Dr. Lijiang Yang again for proofreading this dissertation. Also I thank Dr. Haiyan Wei for the helpful discussions on research projects.

Especially, I want to thank my parents for their unconditional support and encouragement. Without them this goal would not have been achieved.

TABLE OF CONTENTS

CHAPTER Page

LIST OF FIGURES

FIGURE Page

FIGURE Page

FIGURE Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

1.1 The Importance of Proteins in Living Organisms

Proteins which make up more than half of the dry weight of cells are essential parts of organisms.[1] It is commonly believed that creatures would not exist without proteins. Almost all biological functions in the living cell depend on proteins: the translocation of various materials in body fluid is accomplished by protein motors; the locomotion of cells and organisms is controlled by contractile proteins; the receptors for hormones and other signaling molecules are essentially proteins; enzymes which catalyze vital biochemical reactions are proteins.[1] Proteins are basically organic compounds composed of amino acids (residues) which are connected together by chemical bonds between the carboxyl and amino groups of neighboring residues. The sequence of amino acids in a protein is unique and well defined by the sequence of a gene which is encoded in the genetic code.[2] Knowing information of sequence of a protein could only, however, tell us little about its biological functions. In fact, before fulfilling its functions in the living cell, a protein must self-assemble into a particular three-dimensional structure. This self-assembly process is called folding.

One of the fantastic properties of protein is that no matter how complex its sequence is, a protein has the capability to rapidly fold into a characteristic configuration in natural conditions, some as fast as within a millionth of a second. For each protein, the correct three-dimensional folded structure is essential for fulfilling its biological functions.

This dissertation follows the style of *Journal of the American Chemical Society*.

Failing to fold into the correct structure generally leads to the malfunctioning of the protein and consequently to various diseases. For instance, some genetic diseases, such as cystic fibrosis and sickle cell anaemia, are induced by the single residue deletion and mutation respectively.[3] Moreover, Alzheimer's and Parkinson's diseases, spongiform encephalopathies, type II diabetes and prion disease are related to the protein misfolding which generates insoluble protein plaques in the brain and other organs (e.g., heart and spleen).[3,4] Protein plaques normally consist of amyloid fibrils which are characterized by the accumulated cross β-sheet structures with the strands arranged perpendicular to the axis of the fibre.[4] Therefore the studies on proteins, not only on their folding but also on their functions, are of much importance and are among the most popular topics in the field of biology. Experimental and theoretical studies on proteins have been performed incessantly in the last half centenary.

1.2 Methodologies Used in Theoretical Studies of Proteins

Various modern techniques have been applied in experiments to investigate physical and chemical properties of proteins, e.g., the circular dichroism (CD) as well as the laser temperature jumps and time resolved spectroscopy are used to study the folding kinetics of proteins,[5,6] the nuclear magnetic resonance (NMR) and the X-ray crystallography are used to measure protein structures,[7,8] the vibrational circular dichroism (VCD) techniques are applicable for determining protein conformations in the solution.[9] Although experimental studies contribute a lot on the understanding of the folding kinetics and conformational properties of proteins, the latter providing useful structural information of protein functions (e.g., the discovery of multiple conformations of a

protein makes it possible to reveal the transition pathway in its functions), there is still no experimental technique which can provide insight into biological activities at the atomic level. Meanwhile theoretical studies have been developed as the alternative approach of experiments among which computational simulations possess the attractive characteristics of providing the complete picture of biological activities.

In the theoretical studies of biological systems, a variety of simulation methods have been introduced and applied. Among all methods, molecular dynamics (MD) simulation is one of the most popular and matured approaches.[4,10,11] In MD method, the continuous trajectory of the motions of the simulated atomic system (e.g., protein) is generated by integrating the Newtonian equation:[11,12]

$$F_i = m_i \frac{d^2 r_i}{dt^2} \tag{1.1}$$

where F_i is the force on atom i, m_i and r_i are the mass and coordinate of atom i respectively. Several algorithms can be used to integrate the Newtonian equation above numerically,[13-15] all of which are based on the assumption that the coordinate (r) and dynamic properties such as the velocity (v) and the acceleration (a) of atoms could be approximated as Taylor expansions:

$$\begin{aligned} r(t+\Delta t) &= r(t) + \Delta t v(t) + 1/2\Delta t^2 a(t) + 1/6\Delta t^3 b(t) + ... \\ v(t+\Delta t) &= v(t) + \Delta t a(t) + 1/2\Delta t^2 b(t) + ... \\ a(t+\Delta t) &= a(t) + \Delta t b(t) + ... \end{aligned} \tag{1.2}$$

As one of the most efficient algorithms, the leap-frog algorithm uses following relationships:[13]

$$r(t+\Delta t) = r(t) + \Delta t v(t+\Delta t/2)$$
$$v(t+\Delta t/2) = v(t-\Delta t/2) + \Delta t a(t) \quad (1.3)$$

For each atom, its velocity at time $t+\Delta t/2$, $v(t+\Delta t/2)$ is first calculated from its velocity at time $t-\Delta t/2$, $v(t-\Delta t/2)$ and its acceleration at time t, $a(t)$ based on Eq. 1.3. And its position at time $t+\Delta t$, $r(t+\Delta t)$ is obtained as a result of $r(t)$ and $v(t+\Delta t/2)$ that is just calculated. The velocity at time t is achieved by the following relationship:[13]

$$v(t) = \left[v(t+\Delta t/2) + v(t-\Delta t/2)\right]/2 \quad (1.4)$$

In short words, velocities leap frog over coordinates to obtain their values at time $t+\Delta t/2$ and then coordinates leap frog over velocities to obtain their new values at time $t+\Delta t$. Next cycle starts from the velocity $v(t+3\Delta t/2)$ and repeats the leap frog.[12]

One of the important issues in MD simulation is the evaluation of the force F_i in Eq. 1.1. When the simulated system is treated at the fully atomic level, the force is basically the negative derivative of the potential energy. Generally, in MD simulation, interactions between atoms are described as the following potential function:[12]

$$V(r) = \sum_{bonds}\frac{k_i}{2}(l_i - l_{i,0})^2 + \sum_{angles}\frac{k_i}{2}(\theta_i - \theta_{i,0})^2 + \sum_{torsions}\frac{V_n}{2}(1+\cos(n\omega-\gamma))$$
$$+\sum_{i=1}^{N}\sum_{j=i+1}^{N}\left(4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}\right) \quad (1.5)$$

where the first and second terms refer to bond stretching and bond bending interactions between pairs of bonded atoms respectively, the third term is the torsional potential of four bonded atoms and the fourth term corresponds to non-bonded interactions which include electrostatic and van der Waals interactions.[12] This potential energy function is

also called force field. A variety of force fields, which are generally derived from experimental data and quantum mechanics (QM) calculations, have been developed to define suitable parameters in Eq. 1.5 for each type of atom (e.g., CHARMM,[16] AMBER[17] and GROMOS[18] *et al.*). Given the detailed force field, a MD simulation can be performed: the position and velocity of each atom at each step can be calculated and propagated and as a result the complete picture of the overall physical motion of the whole system is finally achieved.

The solvent effect is another important issue in MD simulation since proteins fold and execute their functions in living organisms and therefore the processes are inevitably influenced by the environment (normally solvent). Solvent molecules could be involved in MD simulation either explicitly or implicitly. In an explicit solvent MD simulation, solvent molecules are included in the simulation system and therefore dramatically increase the number of degrees of freedom. Moreover, as a physico-chemical requirement to obtain the accurate simulation trajectory,[19] the step size in integrating the equations of motion (Eq. 1.1) must be a small value (normally set to 1 fs (=10^{-15} s) when explicit solvent model is used). Therefore explicit solvent MD simulation is generally accompanied by enormous computational costs and consequently its application is limited to very small proteins at present. For instance, to simulate the folding of a protein using a PC, whose folding time is 1 ms in reality, suppose that the integration of one step of Newtonian equation takes a CPU time of 0.1 s, the entire calculation time to obtain a fully folded structure of the protein is $0.1 \times 0.001/10^{-15}$ s $\approx$ 3200 year.[19] To reduce this

large computational cost, implicit solvent models have been developed to replace explicit one by largely decreasing the number of degrees of freedom.[11]

The most popular implicit solvent model is the generalized Born/surface area (GB/SA) model,[20,21] in which the solvent is treated as a continuum dielectric media and the solvation effect is quantified by the free energy of transferring a molecule from the vacuum to the solution:[22]

$$\Delta G_{solv} = \Delta G_{pol} + \Delta G_{nonpol} \tag{1.6}$$

where the first term is the free energy change due to electrostatic interactions between the solute and the solvent and the second term is the free energy of transferring the solute into the solution while electrostatic interactions are turned off.

ΔG_{nonpol} is taken to be proportional to the total solvent accessible surface area (SA) of the solute molecule and the value of ΔG_{pol} can be obtained from the solution of Poisson-Boltzmann equation.[23] GB model approaches an approximate numerical solution of Poisson-Boltzmann equation by describing the electrostatic energy due to the interactions between solute and solvent as a sum of pairwise interaction terms between atomic charges:[21]

$$\Delta G_{pol} \approx \Delta G_{gb} = -\frac{1}{2}\sum_{ij} \frac{q_i q_j}{\sqrt{r_{ij}^2 + R_i R_j e\left(\frac{-r_{ij}^2}{4R_i R_j}\right)}}\left(1-\frac{1}{\varepsilon_W}\right) \tag{1.7}$$

where q_i (q_j) is the partial charge of atom i (j), r_{ij} is the atomic distance and ε_W is the dielectric constant of the solvent. R_i (R_j) is the effective Born radius of the interacting atoms i (j) which represents the degree of burial within the solvent for each atom. The

effective Born radius becomes the key parameter for the calculation of solvation free energy and the amount is given by Coulomb field approximation (CFA):[24]

$$R_i^{-1}=\rho_i^{-1}-I \tag{1.8}$$

where ρ_i is the intrinsic radius of atom i and I is the integral over the solute volume of atom i. The integral (I) calculation is the main issue in GB models, with different approaches leading to different versions of GB models. GB models are widely used in the simulation of protein folding and other biological activities. However, the different solutions of I in these models inevitably induce the deviation in results among different models.[25-29] The influence of different implicit solvent models on the folding simulation of proteins will be discussed later in Chapter II.

Besides simplifying the description of the simulated system, another way to reduce the large computational requirements in MD simulation is to develop efficient enhanced sampling algorithms and therefore accelerate the protein configuration sampling. In the last several decades, a variety of sampling algorithms have been developed, e.g., J-walking, adaptive umbrella sampling, multicanonical simulation, metadynamics, conformational space annealing, conformational flooding, potential smoothing methods, hyperdynamics and replica exchange.[30-40] However, there are still difficulties in the application of these algorithms. For instance, the hyperdynamics requires the knowledge of local properties of the potential energy and as a drawback the simulation performed using this method might undersample the low energy range.[34] The J-walking, replica exchange and multicanonical simulation require that simulations must be performed at

different temperatures and therefore they are difficult to achieve continuous simulation trajectories.[41]

In order to improve the efficiency of the sampling and avoid the difficulties of previous methods, we developed two powerful enhanced sampling methods[42,43] and successfully applied them into the folding simulations of several protein systems. The details of the two methods are described in the section of "Simulation Details" in Chapters II and III respectively.

Although MD simulation at the all-atom level can provide more detailed thermodynamic and dynamic information, to study very large biological systems (enormous computational resources are required accordingly), other simulation methods at coarse-grained level are very useful. For instance, multiple coarse-grained Monte Carlo (MC) methods and $G\bar{o}$ -models have been used in the simulations of the folding of large proteins which include the off-lattice MC method,[44,45] the hydrophobic-polar (HP) lattice model,[46,47] the united residues (UNRES) model[48] and *et al.* The coarse-grained normal mode analysis (NMA) (e.g., Gaussian network method (GNM),[49] anisotropic network model (ANM)[50] and chemical network model (CNM)[51] and *et al.*) have been developed for the studies of the global conformational change of proteins. Coarse-grained models generally simplify the structures of polypeptide chains or proteins and consider interactions among groups of atoms rather than detailed atom-atom interactions. Therefore they largely improve the efficiency of the conformation sampling at the expense of ignoring the "fine chemistry" of atomic interactions.[4] Nevertheless coarse-

grained models are still widely used because of their distinct advantage in saving computational costs.

My research includes the developments of efficient methodologies (both all-atom and coarse-grained models) and applications of efficient methodologies into the studies of protein folding and protein functions. At the all-atom level, I employed two enhanced sampling methods developed in the Gao group in implicit solvent MD simulations to investigate the folding mechanism of multiple proteins with β-structures. Meanwhile, at the mesoscopic level, I studied the biological functions of kinesin using a simplified model and its conformational change using the coarse-grained normal mode analysis.

1.3 The Folding Mechanism of β-structures

For many years, protein folding has attracted many scientists' attention. However current studies on protein folding, especially theoretical studies, are still limited to short peptides and protein fragments according to their small size and structural simplicity. Of particular interest are α-helices and β-hairpins.[5,52-56] As two fundamental secondary structures in most proteins, the exploration of the formation of α-helix and β-hairpin is expected to provide an access to the understanding of the folding of global proteins. Experiments display that the formation of β-hairpins is generally complicated than that of α-helices. The latter is normally formed within the time scale of several hundreds of nanoseconds whereas the former is in the time scale of microseconds.[55,57] To date, the folding pathway of individual α-helices and β-hairpins becomes more and more clear but the detailed folding mechanism still remains elusive.

In my studies, I mainly focused on the folding mechanism of β-structures (e.g., β-hairpins and β-sheets). A representative structure of β-hairpin contains two basic structural elements: the turn and strands. Two anti-parallel strands are connected by a turn. Hydrogen bonds are formed between backbone N-H and C=O groups along anti-parallel strands. In addition, cross-strand sidechain interactions exist and play a role in stabilizing the global structure of β-hairpin (see the sample in Fig. 1.1).

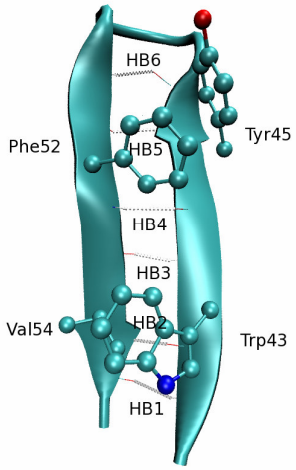


Figure 1.1. A schematic representation of a β-hairpin (the C-terminal hairpin of the B1 domain of protein G). Hydrophobic sidechains of residues Trp43, Tyr45, Phe52 and Val54 are represented by CPK mode. Acceptor-donor pairs for backbone hydrogen bonds are displayed by dash lines.

β-hairpins under studies include not only fragment structures from natural proteins (e.g., those from ubiquitin[58] and human chorionic gonadotropin[59]) but also artificially

designed peptides (e.g., a series of tryptophan zippers (TRPZIP 1 to 6)[6]). The first reported peptide which folds into the native-like β-hairpin structure in water is the C-terminal hairpin of the B1 domain of protein G (GB1 peptide, see the folded structure in Fig. 1.1).[53] Since then the GB1 peptide became one of the most popular hairpin systems, both experimentally and theoretically.[45,53,56,60-64]

The investigation on the GB1 peptide shed light on the folding mechanism of β-hairpin, which displays the cooperativity in the direction along two anti-parallel strands[45,56,61,65] (e.g., in terms of native hydrogen bond formation). The laser-induced temperature-jump experiment performed on the isolated GB1 peptide in solution by Munoz *et al.*[56] suggested a helix-coil model for the probable folding pathway of β-hairpin, in which the folding is initiated from the turn region and then propagates to the terminal along with the formation of backbone hydrogen bonds. The hydrophobic core cluster is packed after the hydrogen bond formation. This model is also called "zipping" or "hydrogen-bond-centric" mechanism and supported by the lattice Monte Carlo (MC) simulation on the folding of the GB1 peptide performed by Kolinski *et al.*[45] Molecular dynamics (MD) simulation on the GB1 peptide by Tsai *et al.*[66] suggested a modification on the zipping model: the folding starts by the turn formation and is followed by the hydrophobic core collapse. The hydrogen bond formation occurs at last. In contrast to the zipping model, another model named "hydrophobic-core-centric mechanism" has been proposed based on several theoretical simulations on the folding of the GB1 peptide.[62,64,67,68] In this model, the hydrophobic core cluster is packed first and simultaneously a portion of backbone hydrogen bonds might be formed. Then the

formation of backbone hydrogen bonds is accomplished and the turn structure is configured. The "hydrophobic-core-centric mechanism" is supported by several recent MD simulations.[69-71]

Which model above is the better description of the folding mechanism of the GB1 peptide? The answer for this question is still obscure. Moreover, a more general question arises on whether there is a universal folding mechanism for all β-hairpins, or in other words, whether the folding mechanism of a given β-hairpin is controlled by its sequence. If its sequence influences the folding mechanism of a hairpin, it would apparently play a role in affecting the formation kinetics and the folding free energy landscape of the β-structure, e.g., changing the formation order of the turn configuration and the hydrophobic core cluster. Previous studies on the sequence influence of β-hairpin folding focused on the folding kinetics and structure stability,[5,61,72-75] whereas a systematic understanding of the sequence influence on the folding mechanism remains to be an interesting research topic. For instance, Munoz *et al.* in their statistical model[73] predicted that the movement of hydrophobic core cluster one residue closer to the turn will largely increase the folding rate of the hairpin. It was also proposed that the function of the hydrophobic core cluster in stabilizing the global structure of hairpin is dependent on its relative position to the turn (the structure is prone to be stabilized only if the hydrophobic core cluster is close enough to the turn).[72,74] Klimov *et al.* by performing the off-lattice model on the GB1 peptide and its mutant suggested that the folding kinetics of hairpin mainly depends on the rigidity of turn.[61]

The series of tryptophan zippers (TRPZIP 1-6) were designed based on the wild-type GB1 peptide and exhibit reversible and highly cooperative thermal unfolding transitions in water.[6] These TRPZIPs differentiate from the GB1 peptide in either the turn sequence and/or the composition of the hydrophobic core cluster in strands. The sequence differentiation and consequent different folding properties of these related hairpins provide an access to understand the sequence influence on the folding mechanism of hairpins. Therefore we use GB1, TRPZIP2, TRPZIP4 and another artificially designed peptide, peptide 1 as samples to simulate their folding. Through the analysis of the free energy landscape as the function of several reaction coordinates corresponding to the three key events involved in β-structure folding, namely the hydrophobic core collapse, the turn formation and the assembling of backbone hydrogen bonds, we prospect to reveal the sequence (represented by the turn sequence and the sidechain hydrophobicity) influence of the folding mechanism of β-hairpins.

In addition to β-hairpins, we also expect to investigate the effects of the turn sequence and the sidechain hydrophobicity on the folding mechanism of β-sheet, which is characterized by a more complex structure composed by multiple anti-parallel strands in comparison to β-hairpins.[76-79] Important factors contributing to the folding and stability of β-sheet structure, similar to those for β-hairpins, include the turn, backbone hydrogen bond as well as sidechain-sidechain interactions across strands.[78,80-83] The three-standed anti-parallel β-sheet, 20mer (also called DPDP), which is characterized by the embedment of unnatural amino acid D-Prolines into two turn segments, was designed by Gellman *et al.*[84] This β-sheet maintains the stable monomeric structure in

the aqueous solution and thereafter has been a popular model system for studies of the β-sheet folding.[79,82,85-87]

In the experimental observation, the stability of 20mer is enhanced with the number of strands increased from two to four.[79] Therefore it was proposed that a three-stranded β-sheet folds via a partially folded structure at which only one hairpin (either N- or C-terminal) is folded and then the formed hairpin works as a template to assist the further folding of the second hairpin.[79,82] A series of mutants (e.g., 20mer^{D}P6D, 20mer^{D}P6D^{D}P14D and 19mer[85,86]) have been designed and both CD and NMR studies were performed for folded structures of these peptides. For instance, the mutation of ^{D}Pro-6 to Asp in 20mer peptide (20mer→20mer^{D}P6D) leads to a change of the turn configuration of the first hairpin in the N-terminal (from type II´ turn promoted by ^{D}Pro to a five-residue TSDGK turn) and the consequent shift in the hydrogen bond assembly.[85] However, it is surprising that the same mutation in the second hairpin doesn't induce any structural change.[86] Among four peptides described above, 20mer and 20mer^{D}P6D are chosen as the samples to run MD simulations to study the folding mechanism of β-sheets. The details of folding simulations of all related peptides are shown in Chapter III.

1.4 Theoretical Studies on the Biological Functions of Kinesin

1.4.1 The asymmetric hand-over-hand mechanism of kinesin walking

In addition to MD simulations of protein folding, my research also includes theoretical studies on the physical motion of kinesin at the mesoscopic scale. Kinesins are microtubule-based protein motors which are involved in many biological functions,

including cargo transportation, mitosis, control of microtubule dynamics, as well as signal transduction.[88-92] As a superfamily, to date, a number of kinesin species have been observed in human being (totally 45 kinesins).[88] Among the different family members, the only conserved domain is the catalytic core.[89] In a general way, kinesins can be classified into three categories depending on the relative position of the catalytic core: N terminal kinesins, C terminal kinesins, and M kinesins, with the first being the majority in human being.[88] The motions of kinesins on microtubules are directional: N terminal kinesins move to the plus end and C terminal kinesins move towards the minus end of microtubule.[88]

The conventional kinesin is a homodimer and each of the monomer contains a heavy chain of ~120 KDa.[92] Each kinesin monomer possesses an N-terminal motor head (the catalytic core) which is responsible for the binding of nucleotides (ATP or the hydrolysis products ADP and Pi) as well as microtubule, a neck-linker, a long coiled-coil which is involved in dimerization, and a globular cargo binding tail domain formed by a light chain (see Fig. 1.2).[88,89,92] Each neck-linker is an extension from the corresponding motor head and is believed to serve as a lever-arm in force generation during the physical motion of the kinesin on the microtubule,[90-92] according to the observed nucleotide-dependent conformational change of the neck-linker.[93-96]

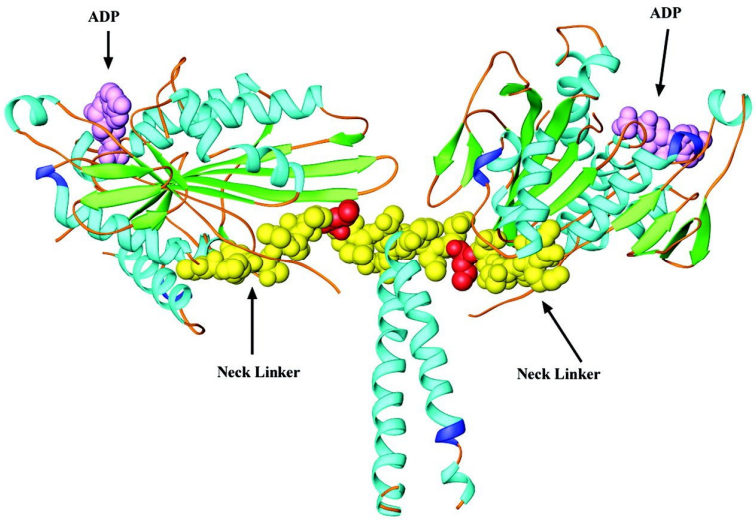


Figure 1.2. A model structure of the kinesin homodimer from *R. norvegicus.*[97] Nucleotides (ADP, the magenta space-filling molecule) and neck-linkers (yellow space-filling residues) are shown in VDW model[*].

As observed in the experiment,[94] the kinesin 1 Cryo-EM exhibits different conformations in the existence of different nucleotides, e.g., ATP binding to a microtubule-associated kinesin head causes a conformational change involving a tilt of the stalk in the forward walking direction (the direction to the plus end of the microtubule). In addition, ADP release induces the redocking of the neck-linker and thus the stepping of kinesin.[98,99] Two distinct states (open and closed) exist in the switch region (switch I) which flanks the active site.[90] Only the closed state is active in ATP hydrolysis, indicating that this switch region is an important element in coupling the

[*] Reprinted with permission notice from "Measuring Kinesin's First Step" by Rosenfeld S. S., Xing J., Jefferson G. M., Cheung H. C., King P. H. *J. Biol. Chem.* 2002, *277*, 36731-36739, Copyright 2002 American Society for Biochemistry and Molecular Biology.

conformational change of kinesin to ATP hydrolysis, also in accordance with the structure difference between the ATP and AMPPNP (an ATP analogue which hydrolyzes very slowly) bound motor domain in fluorescence microscopy studies.[100]

The strength of the interaction between the motor head and the microtubule is also determined by the nucleotide state: detachment force measurement in single molecule experiments confirmed that an ADP occupied kinesin head binds microtubule much more weakly than a nucleotide-free or AMPPNP occupied kinesin head.[101-103] The experimentally determined unbinding force (the force applied to detach kinesin head from microtubule) for the ADP state is smaller than 4 pN,[101,103] whereas the unbinding forces for the empty and AMP-PNP states are both greater than 6 pN.[103] It was also observed that for all three kinds of nucleotide states a larger force is needed to detach the kinesin head in the backword direction (the direction to the minus end of the microtubule) compared to that needed in the forward direction.[103] The unbinding force of the ADP state is 3.3-3.4 pN with a plus-end load and becomes 3.6-3.9 pN under a minus-end load. The unbinding force of the AMPPNP and empty states is 6.1-6.9 pN in the plus direction and is 9.1-10 pN in the minus direction.

The unbinding force is similar for a monomeric and dimeric kinesin if the heads are at the ADP or empty states, indicating that under these conditions only one of the two heads of a dimer binds to the microtubule. The observation that when both kinesin heads are occupied by ADP only one head binds microtubule is consistent with earlier X-ray structural studies.[104] On the other hand, in the presence of AMPPNP, the detaching force is much larger for a dimer than for a monomer, suggesting that under these conditions

both heads are attached to microtubule. It is likely that one of the two heads is occupied by AMPPNP and the other is empty.[103]

The kinetics data for the kinesin have been collected. It has been determined that ATP binds to kinesin with a rate constant of ~4 μM^{-1} s^{-1} and the dissociation constant, $K_{d,}$ is about 75 μM.[90] Therefore, ATP dissociates from its binding site with a rate constant of ~150 s^{-1}.[90] In the absence of microtubule, the rate constant of ATP hydrolysis catalyzed by kinesin is ~6 s^{-1} [90] and ADP releases from a kinesin with a rate constant of ~0.002 s^{-1}.[96] Both ATP hydrolysis and ADP release speed up when kinesin binds microtubule, to 100-300 s^{-1} and ~20 s^{-1}, respectively.[105,106] It was also found that the chemical processes at the two heads are cooperative: the ADP release from one head is further accelerated (60-300 s^{-1}) when ATP is bound at the other head. However, it is believed that ADP release remains the rate limiting step.[90,96] Regardless of the experimental conditions, Pi releases at a rate of >100 s^{-1}.[90] Kinetic measurements also showed that an ADP occupied kinesin head binds the microtubule weakly (K_d~10-20 μM) compared to an empty head.[90]

As observed from single molecular mobility experiments, the conventional kinesin walks along the microtubule with the step size of ~8 nm, the axial distance between two adjacent kinesin binding sites on microtubule.[107-112] This steps size was shown to be invariant in a large range of ATP concentration and external load.[107,109] The stepping of kinesin is tightly coupled to ATP turnover, in the presence of both low and high external load.[107] One step of kinesin requires only one ATP molecule, unless when load is extremely high. Different models have been proposed to explain the processive stepping

of kinesin,[90,92] e.g., an inchworm model was early proposed in which one of the two heads always keeps the leading position.[113,114] However, recent single molecule experiments convincingly showed that kinesin walks by a hand-over-hand mechanism, in which the two heads alternately take the leading position.[112] It has also been observed that kinesin takes not only forward but also backward steps, especially in the presence of external backward load.[115,116] Recently Carter and Cross observed sustained backward steps of an 8-nm step size at large hindering external forces.[116] Backward steps, similar to forward steps, are ATP-dependent. Both forward and backward stepping occur very fast, on the microsecond time scale without detectable substeps, in contrast to some earlier experiments.[108,117]

Based on the experimental observations, Carter and Cross proposed a model in the pre-stroke state of which only one head (presumably at an empty state) binds to microtubule and the other head with an ADP is detached.[116] The detached head takes a position between two microtubule binding sites along the microtubule axial. ATP binding moves the detached head forward and then binds to the microtubule and releases ADP, at the same time the former attached head hydrolyses ATP and detaches from microtubule. This model is more consistent with a stepping pattern with both steps of a kinesin head being intermediate (between 0 and 17 nm and close to 8-nm) but it is in contradiction to the observation that each of the two heads alternately takes steps of ~17-nm and ~0-nm.[112]

Yildiz et al. showed that each kinesin head takes alternative ~17-nm and ~0-nm steps during its walking, with an average step size of the dimer of 8.3-nm.[112] Combined

with an earlier experimental observation that kinesin steps are taken without a stalk rotation,[114] these results suggest that kinesin walks by an asymmetric hand-over-hand mechanism. The asymmetry in the hand-over-hand mechanism was confirmed more recently by several novel experiments in which even the homodimeric kinesin shows 'limping' (every two successive dwell times are separated into longer and shorter ones) under certain conditions.[118,119] In one of the experiments,[119] a single amino acid mutation in the P-loop (nucleotide binding domain) of Drosophila kinesin causes ADP release to be about 3.6-fold faster and the gliding velocity to be 3.3-fold slower. At low forces and/or low ATP concentrations, successive 8-nm steps are observed. However, at high load and high ATP concentrations, one observes 16-nm steps. A careful analysis showed that the 16-nm steps are actually rapid double 8-nm steps. The 16-nm steps are thus due to alternating long and short dwell times. In another experiment, the conventional kinesin, but with truncated stalk, also showed significant limping behavior in the presence of external force.[118] It was shown in this experiment that the extent of limping increases with further shortening of the stalk. Since both experiments mentioned above were performed with homodimeric constructions, the observed limping of kinesin suggests that in the stepping of kinesin, the two heads play different roles (which may not be evident for the conventional kinesin walking under low external load), and the walking is intrinsically asymmetric.

In another experiment, the depletion of the ATP hydrolysis activity of one of the two heads through mutation, so that the ATP hydrolysis by the mutated kinesin head is at least 700 times slower than that by the wild-type one, does not demolish kinesin

walking entirely.[120] The speed of kinesin was indeed reduced by a factor of ~9 as a result of this mutation. This observation is surprising and poses serious challenges to the current understanding of the hand-over-hand mechanism. One of the possible explanations of the above observation is that only one of the two heads plays a dominant role in driving the motion and the other only plays an assisting role. In this picture the two heads are intrinsically non-equivalent, which is inconsistent with the experimental observations that the two heads are chemically equivalent.[90,91,112] The other possibility is that during stepping there exists some kind of rescue mechanism due to the cooperation between the two heads, namely, the mutant head regains some ATP hydrolysis capability when fused with another wild-type kinesin head. On the other hand, an earlier experiment showed that a single-headed kinesin fused with a different protein, which does not have ATP hydrolysis activity but does bind to microtubule, also walks processively along microtubule, although at a slower speed than the conventional kinesin[121]. This experimental observation may suggest that without a possible rescue mechanism, a kinesin with a single motor domain can still function.

Theoretical modeling has proven to be useful in understanding the chemomechanical coupling mechanisms of kinesin. Both kinetic modeling and master equation approaches using thermal ratchet-type models have been used to study the kinetic and mechanical properties of kinesin.[122-126] However, there has been no detailed physical model to describe the external force and ATP concentration dependence in the hand-over-hand mechanism of kinesin walking on a microtubule (in particular, the backward stepping of kinesin) and the asymmetry of kinesin walking. In the present

study, we used the structural and biochemical experimental data to build a physical model to investigate the hand-over-hand mechanism of the kinesin walking on the microtubule. Meanwhile we attempt to unveil the external force and ATP concentration dependence in the walking process of kinesin. Furthermore, the asymmetric property of the hand-over-hand motion of the kinesin was introduced in our model and the reasonable explanation is proposed for the potential origination of the asymmetry. All results are fitted to experimental observations.

1.4.2 The studies on the conformational change of kinesin by using anisotropic network model (ANM)

Two distinct conformations of the monomeric kinesin head have been observed in single molecule experiments:[91-95] in the docked structure, the neck-linker (represented by the green colored segment in Fig. 1.3) is biased to residue 215 which is remote from the nucleotide-binding site but is near the nucleotide ADP (represented by the cyan space-fill in Fig. 1.3 (a)); in the undocked state, the neck-linker is positioned away from residue 215 but is close to the nucleotide binding loop. In the presence of microtubule, kinesin head takes either of the two states as described above depending on the type of the nucleotide attached: the neck-linker exhibits strong preference for the conformation in the docked state with the existence of ATP or its analogue AMPPNP whereas for the conformation in the undocked state with ADP occupied or nucleotide-free.[95] The same conformations of the neck-linker were also observed in the dimeric kinesin and therefore the neck-linkers are believed to serve as lever-arms during the kinesin walking.[95]

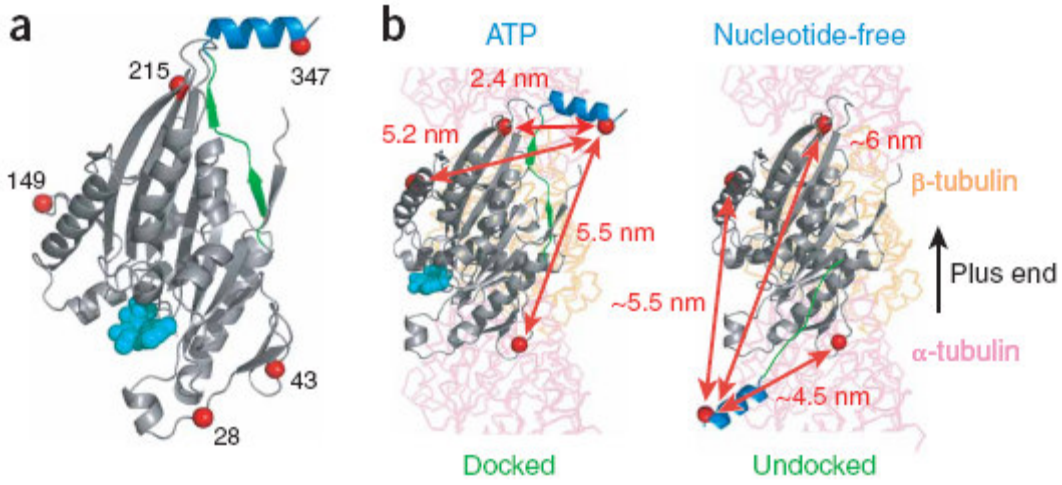


Figure 1.3. Single molecule measurement of monomeric kinesin.[95] (a) the crystal structure of rat kinesin monomer (PDB: 2KIN[127]) with the neck-linker (green) and coiled-coil (blue) in the docked state. ADP is represented by the cyan space-fill. (b) Schematics showing the docked and undocked structures of the monomeric kinesin in the presence of microtubule.*

The global conformational change of the kinesin head induced by the binding of the nucleotide as described above is essentially the collective motions of subdomains. The global motion generally possesses the lower frequency in comparison to local fluctuations of individual atoms. The two categories of motions (the local fluctuation of atoms and global conformational change of the whole protein) consist of the main

* Reprinted by permission from Macmillan Publishers Ltd: *Nat. Struct. Mol. Biol.,* "Single-molecule Observations of Neck-linker Conformational Changes in the Kinesin Motor Protein" by Tomishige M., Stuurman N., and Vale R., 2006, 13, 887-894, Copyright 2006 (http://www.nature.com/nsmb).

contents of the internal dynamics of proteins and become a subject of considerable interest.[128,129] One of the principal tools in the theoretical study of internal dynamic properties of proteins is MD simulation, the methodology of which is described earlier.

An alternative approach to elucidate internal motions of proteins is the normal mode analysis (NMA). The standard NMA is based on a principal assumption that the potential energy of a global protein (E_p) could be approximated as a quadratic function of the generalized coordinates q_i in the neighborhood of a defined energy minimum at the equilibrium state q_i^0 :[128,130,131]

$$E_p = \frac{1}{2}\sum_{i,j} F_{ij}(q_i - q_i^0)(q_j - q_j^0) \tag{1.9}$$

where $F_{ij} = \frac{\partial^2 E_p}{\partial q_i \partial q_j}|_{q=0}$ is the generalized force matrix (the second derivative function of the potential energy). The insertion of the quadratic function of the potential energy as well as the kinetic energy (E_k) into Lagrangian equation ($\frac{d}{dt}(\frac{\partial L}{\partial \dot{q}_i}) = (\frac{\partial L}{\partial q_i})$, where $L = E_k - E_p$) eventually generates a set of "normal modes", which are essentially vibrational modes. The overall motion of atoms is therefore expressed as the linear combination of independent normal modes. For a protein system with N atoms, (3N-6) normal modes are finally obtained from the eigen-analysis of the second derivative function of the potential energy (F_{ij}). With knowledge of normal modes, correlation functions of atomic fluctuations and crystallographic temperature factors (B-factors)

could be calculated. Numerous examples showed that experimentally measured B-factors are well reproduced by NMA.[49,128,130,131]

Similar to MD simulation, the application of the standard all-atom NMA is also limited to small biological systems due to its large computation requirements. Recently a coarse-grained NMA, anisotropic network model (ANM), was developed to study the protein dynamics based on C_α atom coordinates.[49,50,132-134] In this model,[50] a global protein of interest is described as an elastic network and C_α atoms of all residues are regarded as nodes of the network. All nodes are connected by harmonic springs which are set to be uniform in the study of Tirion *et al.*[49] (all springs have the same force constant). Without considering the detailed feature of the potential energy, such as the details of chemical bond stretching energy, bond angle bending energy and torsional angle twisting energy as described in Eq. 1.5, the overall potential energy in this model is described by a harmonic function of the inter-node distance. However, in spite of the shortage of the detailed residue information and the consequent roughness of the description of interactions between atoms (the interactions might be anharmonic instead of harmonic in the real environment), ANM plays an important role in the dynamics investigation of proteins because of its ability to provide insight into the large-scale conformational changes of proteins.[135-138] The possible explanation might be that the large-scale slow motion is the consequence of the collective motions of subgroups of atoms.[49] The sum of inter-atomic interactions eventually approaches a universal form, in the conformity with the central limit theorem. Therefore the details of atomic interactions could be ignored compared to the net slow motion.[49]

Numerous studies have shown that the functionally important global motions of proteins could be well represented by single or sometimes the combination of a few low-frequency normal modes.[139,140] A large degree of overlap between low-frequency normal modes predicted by ANM and the experimentally observed displacement vector between known conformations was obtained in many biological systems.[50,138,140-142] In most cases, the most functionally relevant mode is not the lowest-frequency one.[138,140,141] Therefore it is interesting to explore the "conformational changes" simulated by low-frequency normal modes and evaluate the similarity between each mode and the real function of the protein. More recently, based on the Gaussian network model (GNM),[49] another coarse-grained model called chemical network model (CNM) was proposed in order to emphasize the chemical information of residue interactions in the elastic network.[51] The inter-node interaction therein is determined by the closest nonhydrogen atoms of two residues rather than C_α atoms. The types of interactions are then separated into several classes (e.g., polar, nonpolar *et al.*) depending on the relevant atom types.[51] This method further improves the accuracy on the prediction of B-factors.

As described earlier, the structure and dynamics of a protein is inevitably influenced in the presence of solvent molecules.[143,144] The solvent is treated explicitly[145-148] or as a continuum dielectric region in MD simulation.[20,21] In coarse-grained models the collisions and friction forces are introduced to describe the collisions between the solute and solvent molecules.[11] The dynamic effect appears in terms of frictional as well as stochastic forces if the protein molecule is treated as a Brownian particle in solution, as described in Langevin equation.[144,149,150] As a direct result, the interactions of solvent

molecules on the protein energy surface exert a damping effect on all vibrational motions. Langevin mode analysis was developed by Lamm and Szabo in 1985[144] to describe the behavior of atoms, which move on a harmonic potential surface, under the damping effect described by the Langevin equation. The details of Langevin mode analysis are described in the section of "Theoretical Models" in Chapter V.

In summary, we used ANM method to study the motions of kinesin and then used Langevin mode analysis to study the damping effect on each normal modes yielded in ANM. As described above, low-frequency normal modes make dominant contributions to the large-scale conformational changes of proteins, therefore the Langevin mode analysis mainly focused on the functionally important low-frequency normal modes.

CHAPTER II

A TEST OF IMPLICIT SOLVENT MODELS ON THE FOLDING SIMULATION OF β-HAIRPIN*

2.1 Introduction

The generalized Born/surface area (GB/SA) model has been applied widely in MD simulation of protein folding. As an implicit solvent model, the solvent in GB/SA model is treated as a continuum dielectric region.[20,21] Poisson-Boltzmann equation can be used to solve the solvation energy and GB model provides the numerical solution of Poisson-Boltzmann equation (see Eq. 1.7). The effective Born radius of each solute atom (R_i) becomes the key element for the solvation energy calculation and the integral over the solute volume of each atom (I) determines the amount of the effective Born radius of the corresponding atom[24] (see Eq.1.8 in Chapter I). Different solutions of the integral I lead to different versions of GB models (e.g., GB^{HCT}, GB^{OBC} and GBn models).

In the early GB^{HCT} model (igb=1),[151] the integral I is estimated over van der Waals (VDW) sphere of each protein atom, which, as a drawback, creates regions of interstitial high dielectrics where the solvent molecule is too large to enter.[22] As a result, the effective Born radius for buried atoms are underestimated.[22] To reduce the influence of interstitial high dielectrics, in GB^{OBC} model (igb=2 or 5), the effective Born radius is given by a well-behaved three-parameter scaling function (α, β and γ) with the integral derived from HCT approach:[152]

---

*Reprinted with permission notice from "A Test of Implicit Solvent Models on the Folding Simulation of the GB1 Peptide" by Shao Q., Yang L. J., Gao Y. Q., *J. Phys. Chem.* 2009, 130, 195104/1-195104/6, Copyright 2009 American Institute of Physics.

$$R_i = (\tilde{\rho}_i^{-1} - \rho_i^{-1} \tanh(\alpha I\tilde{\rho} - \beta(I\tilde{\rho})^2 + \gamma(I\tilde{\rho})^3))^{-1} \quad (2.1)$$

where $\tilde{\rho}_i = \rho_i - 0.09$ Å , ρ_i is the VDW radius of atom i and α, β and γ are tunable parameters. In GBn model (igb=7),[22] a correction term is first added to the integral over the VDW surface to make up the difference between VDW and molecular surfaces, then the scaling function is performed with the new integral and a new set of α, β and γ values to achieve the accurate effective Born radius.[22] As a result, GB^{OBC} model produces geometry-independent correct effective Born radius whereas the calculation of effective Born radius in GBn model is geometry-specific.[22]

One question regarding GB/SA models is how well the models reproduce or predict the kinetics and thermodynamics of protein folding. For the sake of answering this question, Zhou performed REMD simulations on the folding of GB1 peptide with different force fields in combination with different implicit models and compared the results to that from explicit solvent model.[29] Of the implicit solvent models studied, only AMBER96/GBSA reproduced reasonable results comparable to the explicit model. More recently Shell *et al*. tested the stability of several peptides (GB1 peptide, TRPZIP2, C peptide and EK helix) with AMBER force fields and GB/SA models with different generalized Boltzmann models.[28] It turned out that the combination of AMBER96 with GB^{OBC} model (igb=5) is the best choice to balance the α-helix and β-hairpin tendencies of the peptides tested.[28]

In the present study, we used an enhanced sampling method[42] to investigate the folding behavior and the relevant free energy surface of GB1 peptide[27] (sequence: GEWTYDDATKTFTVTE, PDB code: 2GB1,[153] see the folded structure in Fig.1.1)

quantitatively. The two best behaved models from the studies of Shell *et al.*[28] (GB^{OBC} and GBn) were tested to study quantitatively the influence of implicit models on the folding mechanism of GB1 peptide. Multiple folding and unfolding transitions between the folded and the extended conformation were observed in several independent MD trajectories, each running in the time scale of several hundred nanoseconds. It is worth mentioning that the dynamics information in the enhanced sampling simulation is lost due to the exertion of the biased potential energy. Therefore the present study mainly focused on the thermodynamics of the β-hairpin folding.

2.2 Simulation Details

Molecular dynamics simulation provides direct information on the detailed dynamic folding events at the atomic level. However, even with today's computer powers, computational simulation studies on protein folding still have to be confronted with a challenging problem, the huge computational cost which is induced by the complex potential energy surfaces involved in protein folding. Potential energy surfaces of complex systems might possess very complicated features because of the existence of many local energy minima and barriers. The energy (conformation) sampling in the standard MD simulation on the potential energy surface therefore could be trapped in local minima easily and herein take long time to cross the energy barrier. Here in the enhanced MD simulation,[42] we simply add a bias potential to the potential energy surface of the real system to enhance the sampling of the energy range which is sparsely sampled in standard MD, e.g., the high energy range. As a result, the sampling over the energy range which is heavily sampled originally is decreased and the sampled energy

distributions become more uniform, keeping all the important energy states involved. The bias potential function herein is provided by a sum of Gaussian functions:[42,154]

$$f(V) = \sum_{i}^{n} a_i e^{-(\frac{V-V_i}{\sigma_i})^2} \tag{2.2}$$

where V_i is the energy of a state which is prone to be sampled with enhancement and a_i is a negative constant defining how much the enhancement will be.

In order to determine the values of parameters in Eq. 2.2, the detailed information of the original potential from the standard MD is required. Therefore, firstly, short trajectories are run with the standard MD simulation at the desired temperature from the native structure and the fully extended structure of the protein respectively. The two trajectories last normally from several hundreds of peco-seconds to 1 ns and then are analyzed to obtain enough information of the energy distribution which covers the transition between folded and unfolded conformations of the protein. The parameters in Eq. 2.2 are determined in such a way that the magnitude of *f(V)* is small for lower energy ranges in the standard MD simulation, but large for higher energy ranges. As a result, the sampling over both lower and higher energies is enhanced. With knowledge of the certain parameter values, the enhanced MD simulation is performed in long time and received trajectories are used to study thermodynamic properties of the protein by using standard recovering procedures.

The enhanced MD simulation with the biased potential $\tilde{V}(r) = V + f(V)$ yields a distribution function over the configuration space:

$$\tilde{\rho}(r) = e^{-\beta \tilde{V}(r)} / \tilde{Q} \tag{2.3}$$

where $\beta = 1/k_B T$ and $\tilde{Q} = \int e^{-\beta \tilde{V}(r)} dr$ is the partition function. The distribution function with the original potential thus can be recovered from the biased one as:

$$\rho(r) = e^{-\beta V} / Q = e^{-\beta \tilde{V}} \times e^{-\beta (V - \tilde{V})} / Q = \tilde{\rho}(r) \times e^{-\beta (V - \tilde{V})} \tilde{Q} / Q \tag{2.4}$$

where $Q = \int e^{-\beta V(r)} dr$.

With the enhanced MD method the simulation was performed by using AMBER 9.0 package. The peptide was modeled with the AMBER FF96 all-atom force field[155] and the two GB implicit solvent models.[20,21] The salt concentration is set to 0.2 M and the default surface tension is 0.005kcal/mol/Å^2. The SHAKE algorithm[156] with a relative geometric tolerance of 10^{-5} is used to constrain all chemical bonds. The simulation was carried out at the room temperature which was maintained by using the weak-coupling algorithm[157] with a coupling constant of 5.0 ps^{-1}. No non-bonded cutoff was used in the simulation.

The native-like or namely folded structure of GB1 peptide is defined by following criteria: 1) the hydrophobic core is well packed (the core radius of gyration R_g^{core} is less than ~3.7 Å), 2) the all-residue root mean square deviation (RMSD) from the GB1 NMR structure is less than 2.5 Å ($RMSD = \sqrt{\frac{1}{N}\sum_i^N (r_i - r_i^0)^2}$), 3) at least four of six native hydrogen bonds are formed (the hydrogen bond inside of turn, D46-T49 is not included because of its instability). Here we name hydrogen bonds HB 1-6 ordered from the tail to the turn (see Fig. 1.1). A hydrogen bond is formed only if the distance between the carbonyl oxygen and amide hydrogen (O...H) is less than 3.5 Å and the NHO angle is greater than 145^o.

2.3 Results

A total of 11 independent trajectories were obtained with each one starting from a fully extended conformation and running in the time scale of hundreds of nanoseconds. The total simulation time is up to ~2μs. Two typical trajectories obtained from different implicit models are shown in Fig. 2.1 represented by the time series of all-residue RMSD. One can see clearly from the figure that GB^{OBC} model is better at yielding structures with small RMSD compared to GBn model.

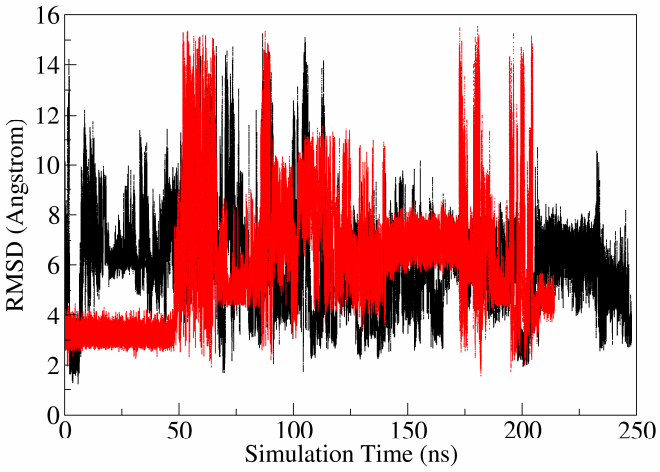


Figure 2.1. Time series of heavy-atom RMSD value in the two typical trajectories from the enhanced MD simulations (black line representing the case of GB^{OBC} model, red line representing the case of GBn model).

Figure 2.2 illuminates the free energy landscape as a function of the radius of gyration of the hydrophobic core R_g^{core} and the number of native hydrogen bonds N_{HB} in the models of GB^{OBC} and GBn.

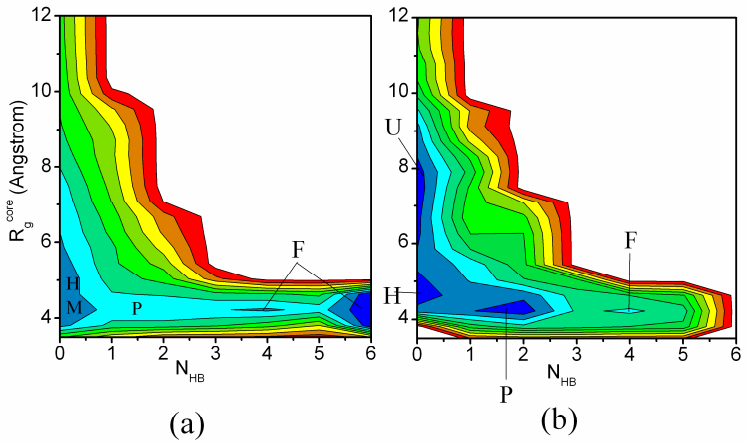


Figure 2.2. The free energy landscape as a function of R_g^{core} and N_{HB} for (a) GB^{OBC} model and (b) GBn model. The contours are spaced at intervals of $2k_BT$.

Similar features can be seen from the two free energy maps: 1) Both maps show an "L" shape, which reveals that the early stage of folding process is driven by hydrophobic core collapse, not by hydrogen bond interactions. The latter works as the driving force for the folding in the later stage, meanwhile the compact hydrophobic core cluster plays an important role in stabilizing the global structure; 2) Several populated states are presented in the folding pathway of GB1 peptide: the unfolded state (U), the molten globule (H, R_g^{core} ~5.0 Å and $N_{HB} \leq 1$), the partially folded (P, R_g^{core} ~3.5 Å and $2 \leq N_{HB}$

$\leq$3) and the folded (F, $N_{HB} \geq 4$) states. Clustering analysis gives average structures of these populated states, as shown in Fig. 2.3 (a). Moreover, multiple misfolded states (M) are observed from the clustering analysis, as shown in Fig. 2.3 (b). These states are characterized by a symmetric β-hairpin structure but with the totally opposite orientation of strands (R_g^{core} ~3.5 Å, C_α_RMSD<2.0 Å), one-residue shifted hairpin structures, either to the N-terminal or to the C-terminal (R_g^{core} ~6.0 Å, C_α_RMSD<2.6 Å). All misfolded states are stabilized by multiple non-native backbone hydrogen bonds whereas none of the six native hydrogen bonds is fored.

In summary, major differences exist in the free energy profile obtained using the two different implicit solvent models (Fig. 2.2): 1) Most importantly, the lowest energy state in GB^{OBC} model is the native state (N_{HB}=6), whereas in GBn model it is H or P state. This indicates that compared to GBn model, GB^{OBC} model is much better in capturing the folded state of β-hairpin. 2) Misfolded states, as described above, are presented in the free energy map of GB^{OBC} model. No apparent energy barrier is observed between misfolded states and H state. 3) The P state in GB^{OBC} model has a significantly higher free energy than that in GBn model.

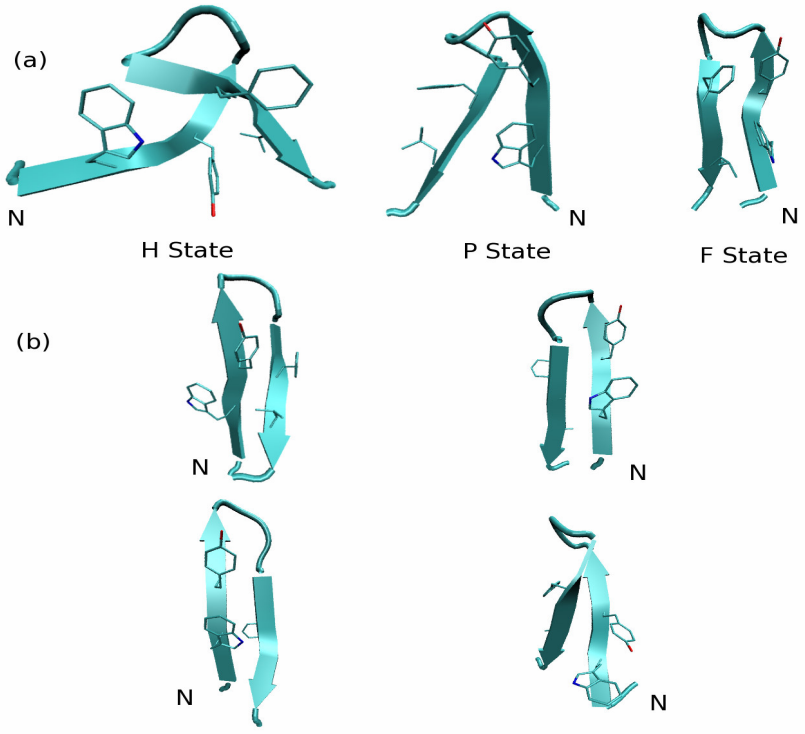


Figure 2.3. Representative structures of (a) H, P and F states, (b) misfolded states from the clustering analysis in GB^{OBC} model. The hydrophobic core is shown in licorice mode. (b) left upper: symmetric hairpin structure with four non-native hydrogen bonds (43:HN-54:O, 45:HN-52:O, 52:HN-45:O, 54:HN-43:O), (b) right upper: shifted hairpin structure with four non-native hydrogen bonds (42:HN-54:O, 44:HN-52:O, 52:HN-44:O, 54:HN-42:O), (b) left lower: shifted hairpin structure with five non-native hydrogen bonds (43:HN-53:O, 45:HN-51:O, 51:HN-45:O, 53:HN-43:O, 55:HN-41:O), (b) right lower: shifted hairpin structure with four non-native hydrogen bonds (44:HN-54:O, 46:HN-52:O, 52:HN-46:O, 54:HN-44:O)

In order to compare the performance of the two implicit solvent models, explicit solvent simulations using the AMBER ff96/SPC force field were conducted and the corresponding free energy landscape (as shown in Fig. 2.4) was used as a reference. By comparing the two implicit solvent models, GB^{OBC} is shown to yield a free energy landscape in better agreement with the explicit solvent model of AMBER ff96/SPC force field: both show an "L" shape free energy surface in which the native state is the lowest free energy state. Besides, the free energy difference between H and F state is about 0.90 k_BT in the simulations using AMBER ff96/GB^{OBC}, which is in good agreement with the value of 1.07 k_BT in the simulations using AMBER ff96/SPC.

However, different solvent models do introduce differences in the populated states. For instance, the native state F of GB1 peptide is characterized by R_g^{core} <5 Å and N_{HB}=5 and the H state is characterized by R_g^{core} ~3.7 Å and $N_{HB}\leq 1$ in the explicit solvent simulations; while the F state is characterized by R_g^{core} <5 Å and N_{HB}=6 and the H state is characterized by R_g^{core} ~5 Å and $N_{HB}\leq 1$ in the simulation results obtained using the GB^{OBC} implicit solvent model. In addition, the low-lying regions in the free energy landscapes of the implicit solvent simulations (Fig. 2.2) are narrowly centered at the free energy minimum for the state F, while the same region spreads more widely in the free energy landscape obtained from the explicit solvent simulations (Fig. 2.4).

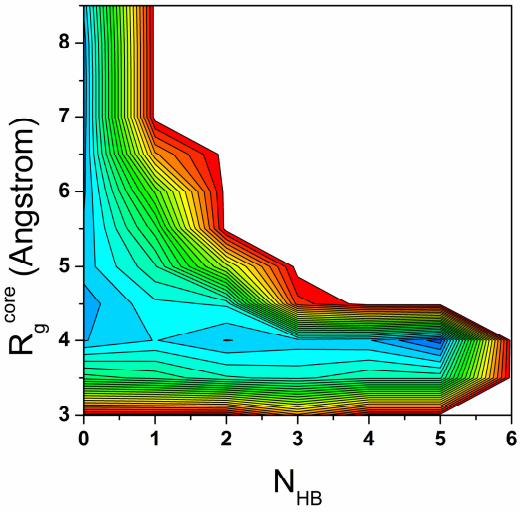


Figure 2.4. The free energy landscape as a function of R_g^{core} and N_{HB} for the explicit solvent of AMBER ff96/SPC. The contours are spaced at intervals of $0.5k_BT$.

As an alternative representation, we calculated the free energy landscape as a function of R_g^{core} and C_α_RMSD and as a function of N_{HB} and C_α_RMSD as shown in Fig. 2.5. From Figure 2.5 (a) (GB^{OBC} model) one can see a single energetic descent to the folded state. In contrast, the simulation using the GBn model generates a much more complex free energy landscape (see Fig. 2.5 (b)). Several stable intermediates exist during the collapse of the hydrophobic core. Figure 2.5 (c) & (d) illuminate the free energy map as a function of N_{HB} and C_α_RMSD in both implicit models. The native state has the lowest free energy in GB^{OBC} but not in GBn model. In summary, GB^{OBC} model is better at yielding folded state of GB1 peptide.

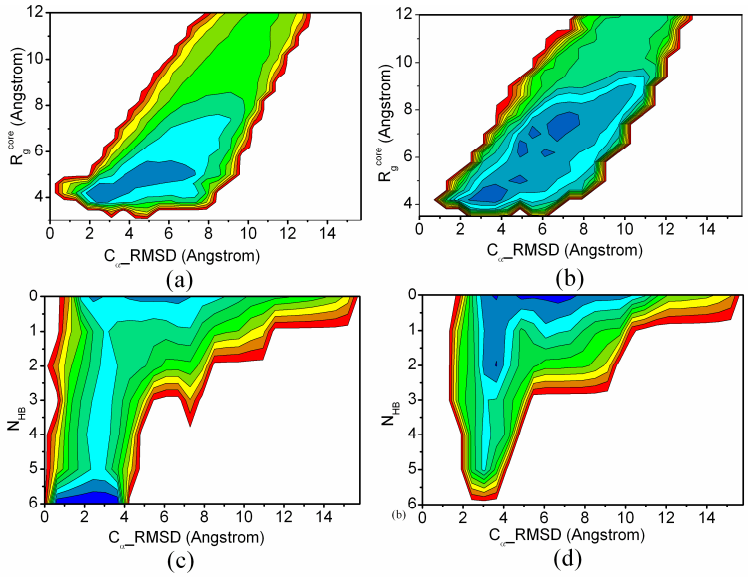


Figure 2.5. The free energy landscape as a function of R_g^{core} and C_α-RMSD for (a) GB^{OBC} model and (b) GBn model; the free energy landscape as a function of N_{HB} and C_α-RMSD for (c) GB^{OBC} model and (d) GBn model. The contours are spaced at intervals of $2k_BT$.

From the free energy landscape (Fig. 2.2 (a) and Fig. 2.5 (c)) obtained from GB^{OBC} model, we can see that the folding of GB1 peptide is not a single downhill process. Indeed there exist misfolded states and several intermediates between the unfolded and folded states. The ultrafast initial structure collapse drives the transition from U to H state. H state then could take the path to either folded (F) state or misfolded (M) state. For the transition from H to F state, the free energies of H and P states are almost similar. Considering the observation of low population of P state in the free energy profile in GB^{OBC} model, it is reasonable to regard the folding of GB1 peptide as a two-state transition with H and F states as the main minima.

Several questions regarding the folding of GB1 peptide are still controversial, e.g., the formation order of backbone hydrogen bonds.[9, 18, 20] The hydrogen bond formation could start from the inner and end at the terminal (hydrogen bond-centric "zip-out" folding mechanism) or in reverse ("zip-in" mechanism). The "zip-out" mechanism was suggested as the more probable one.[56] To investigate the formation order of hydrogen bonds in GB1 peptide, we performed an analysis for the transition state ensemble. Firstly, the transition state ensemble was defined by R_g^{core} <5.0 Å, C_α_RMSD >3.5 Å and N_{HB}=2 or 3. Then following the criteria above, the suitable structures were chosen from all trajectories. Through the clustering analysis, the chosen structures are mainly separated into two clusters. The major cluster which contains 94.0% of transition state ensemble has a typical structure in which the hydrophobic core is packed and the nearby hydrogen bonds HB 4 and 5 are formed. In contrast, HB 3 is sparsely formed and HB 1 and 2 near the terminal are never formed in the structures of this cluster. On the other hand, in the minor cluster (6.0%), only HB1 and 2 are formed. Surprisingly HB 6 closest to the turn is never formed in all snapshots of the selected transition state ensemble.

By analyzing all successful folding trajectories we observed that the formation and stabilization of HB 6 mostly occur in the end of the assembly of hydrogen bonds. In several folding events, although HB 6 is formed early following the hydrophobic core collapse, the bond is weak and breaks quickly. Moreover, the analysis of the accelerated folding trajectories (although the real dynamics information is lost in these trajectories) demonstrated that in most successful folding events (18 out of 19) HB 5 forms first and remains stable through the entire folding process and then the rest of hydrogen bonds are

formed following the order from HB4 to HB1. Only one folding event starts the assembly of hydrogen bonds from HB1 and 2 near the terminal. These results may suggest that the hydrogen bond formation mainly follows the "zip-out" mechanism: the hydrogen bond formation starts from the position near the turn and propagates to the end.

2.4 Conclusion

In this chapter we investigated the influence of implicit solvent models on the folding simulation of GB1 peptide. We calculated the free energy landscape for the folding of GB1 peptide using AMBER ff96 force field with two different implicit solvent models, namely GB^{OBC} and GBn. The two implicit models are indeed the correction of the early GB^{HCT} model[151] to eliminate the influence of the regions of high dielectrics, the existence of which leads to the underestimate of the effective Born radius especially for the buried atoms.[22] GB^{OBC} uses a three-parameter rescaling function to rescale the effective Born radius with the integral of VDW sphere of each atom firstly derived from GB^{HCT} model. GBn corrects the integral by adding a geometry-dependent molecular volume correction term and then uses OBC methodology to determine the effective Born radius.

The simulation results are compared to those obtained from the explicit solvent simulations. In contrast to GBn model, the combination of ff96 force filed and GB^{OBC} model performs better in simulating the folding of GB1 peptide. It is well known that the commonly used force fields in AMBER (ff94, ff96, ff99) have different propensities to form secondary structures because of the differences in backbone torsion energies.

Zaman et al. claimed that ff96 force field strongly favors the hairpin conformations whereas ff94 and ff99 favor the helical conformations.[158] The torsion potential of AMBER was modified by Gnanakaran and coworker to obtain better agreement with experiments for the helix-coil transition.[159] However, in the more recent work by Ozkan et al., it was found that ff96 force field is better balanced for various secondary structures than the other force fields.[160] In our simulations, it is apparent that when the ff96 protein force field is used the GB^{OBC} implicit solvent model is a more suitable choice than the GBn model for the modeling of the β-hairpin since the former but not the latter yields folding free energy profiles similar to those obtained in the explicit solvent simulations. Certainly, more studies are needed to evaluate how well the implicit solvent simulation will represent the explicit solvent for a general protein. But that is beyond the scope of the present study and we mainly focus on the influence of different implicit solvent models on β-hairpin folding.

The thorough analysis of the free energy landscape as a function of Rg^{core} and N_{HB} demonstrates that the folding of GB1 peptide can be treated as a two-state transition with the local minima of molten globule (H) and the folded (F) states (see Fig. 2.2). Both hydrophobic core cluster and cross-strand hydrogen bonds play important roles in the hairpin formation. The clustering analysis of the transition state ensemble and the hydrogen bonds support a mechanism in which the hydrogen bond formation mainly obeys the "zip-out" mechanism. The formation of HB 6, the hydrogen bond closest to the turn structure, turns out to be difficult. One reasonable explanation for the difficult formation of HB 6 is that it is linked to the difficult formation of the turn configuration,

which occurs in the last stage of GB1 peptide folding. On account of its nice behavior on the simulation of GB1 peptide, GB^{OBC} implicit model, in combination with AMBER force fields, is prospected to have more applications in simulating peptides and small proteins.

CHAPTER III

EFFECTS OF SIDECHAIN HYDROPHOBICITY AND TURN SEQUENCE ON THE FOLDING OF β-STRUCTURES*

3.1 Introduction

In Chapter II we demonstrated the better behavior of GB^{OBC} in the folding simulation of GB1 peptide compared to GBn model. The simulation results also indicated that the folding of GB1 peptide is a two-state transition. Starting from the fully extended structure, the folding of GB1 peptide is driven by the fast hydrophobic core collapse to H state. The transition from H state to the fully folded (F) state is driven by the assembling of backbone hydrogen bonds. However, the information related to the turn structure formation was not provided.

Previous experimental and theoretical studies on the GB1 peptide suggested two models explaining the mechanism of the β-hairpin folding, "hydrogen-bond-centric mechanism"[56,73] and "hydrophobic-core-centric mechanism".[62,64,68] Moreover, Tsai *et al.* suggested a modification on the former model.[66] The details of these models are described in Chapter I. The contradiction displayed among these models focuses on the formation order of the three key elements, the hydrophobic core cluster, the turn and the backbone hydrogen bond assembly.

Therefore, a deeper analysis on the results of our folding simulation of GB1 peptide, especially on the turn formation, is necessary to be carried out in order to reveal which model above bestly describes the folding process of GB1 peptide. Furthermore, it is

---

important to understand whether the model for the folding of GB1 peptide is also suitable for the folding of all β-hairpins. If a universal folding mechanism exists for all β-hairpins, the folding of a hairpin must follow the exact process no matter what its sequence is. Otherwise the folding mechanism of a β-hairpin should be determined by its sequence. To understand the sequence influence on the folding mechanism of hairpins we used AMBER ff96 force field and GB^{OBC} model to simulate the folding of several related peptides besides GB1 peptide, e.g., peptide 1 (sequence: $SESYIN^{D}PDGTWTVTE$), TRPZIP2 (sequence: SWTWENGKWTWK, PDB code: $1LE1^{6}$)[161] and TRPZIP4 (sequence: GEWTWDDATKTWTWTE, PDB code: $1LE3^{6}$) (see the folded structures in Fig. 3.1). The integrated tempering sampling (ITS) method was used to enhance the energy sampling in the simulation.

Sequences of these four hairpins differentiate in two aspects: the hydrophobic core cluster and the turn sequence, which are believed to be the key elements for the kinetics and thermodynamics of the hairpin folding. Peptide 1 possesses the weakest hydrophobic interactions (mainly between residues Ile5 and Trp11) and has a strongly favored type I´ turn in the folded state. TRPZIP4 is a mutant of the natural protein fragment GB1 peptide, with three of the four hydrophobic residues (Tyr5, Phe12 and Val14) replaced by tryptophans. TRPZIP4 and GB1 possess the same turn sequence (a type I turn formed by 6 residues) and the same cross-strand native hydrogen bond assembly. TRPZIP2, a 12-residue peptide, possesses the same hydrophobic core cluster as TRPZIP4 whereas a different 4-residue type I´ turn.

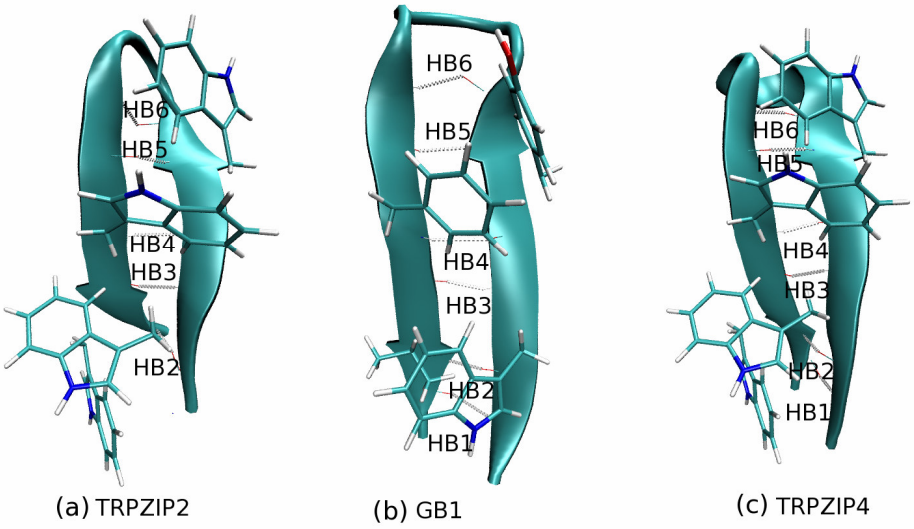


Figure 3.1. The folded structure of (a) TRPZIP2, (b) GB1 and (c) TRPZIP4. The hydrophobic core is shown in licorice mode and backbone hydrogen bonds are shown in dash lines.

In summary, peptide 1 and TRPZIP2 have more strongly favored turn structures compared to GB1 and TRPZIP4. On the other hand, TRPZIP2 and TRPZIP4 have very strong hydrophobic interactions (accordingly, large sidechain hydrophobicity) whereas the hydrophobic core stability of GB1 is weaker and that of peptide 1 is weakest.

In addition to hairpins, two three-stranded β-sheets were also simulated using ITS method in order to further understand the sequence influence on the folding mechanisms of β-structures, which are 20mer (sequence: VFITS^{D}PGKTYTEV^{D}PGOKILQ) and its mutant 20mer^{D}P6D (sequence: VFITSDGKTYTEV^{D}PGOKILQ). For all β-structures, the relative stability of backbone hydrogen bonds was studied and herein its dependence on the turn sequence and the sidechain hydrophobicity was predicted.

3.2 Simulation Details

The ITS approach applied in MD simulation has been described previously.[43,162] Briefly, the generalized distribution function, as a function of the potential energy U, can be written as an integration over β:

$$p(U) = \int_{\beta} f(\beta') e^{-\beta' U} d\beta' \tag{3.1}$$

where $\beta = \frac{1}{k_B T}$ (k_B is the Boltzmann constant and T the temperature) and $f(\beta')$ is a function containing information of the temperature dependence of the partition function. In MD simulation on a modified potential U´ (as a function of the original potential U) at the desired temperature corresponding to β, the distribution function could be rewritten as:

$$e^{-\beta U'} = p(U) = \int_{\beta'} f(\beta') e^{-\beta' U} d\beta \tag{3.2}$$

The modified potential is:

$$U' = -\frac{1}{\beta} \int_{\beta'} f(\beta') e^{-\beta' U} d\beta' \tag{3.3}$$

The function $f(\beta')$ is estimated in ITS by a quick and robust procedure in order to achieve an even sampling in the desired energy range[43] and the biased force in MD simulation is then obtained as the derivative function of U´ in Eq. 3.3. Thermodynamic properties of the system are finally calculated by reweighting corresponding terms with a weighting factor of $e^{\beta_0 (U' - U)} = e^{-\beta_0 U} / p(U)$.

MD simulation was performed by using AMBER 9.0 package. All peptides were modeled with AMBER FF96 all-atom force field[155] and GB^{OBC} implicit solvent model.[20,21] The salt concentration is set to 0.2 M and the default surface tension is 0.005kcal/mol/Å^2. The SHAKE algorithm[156] with a relative geometric tolerance of 10^{-5} is used to constrain all chemical bonds. The simulations were carried out at the room temperature which was maintained by using the weak-coupling algorithm[157] with a coupling constant of 5.0 ps^{-1}. No non-bonded cutoff was used in the simulation.

3.3 Results

3.3.1 Criteria for the folded state of β-hairpins and β-sheets

Three criteria were used to define the folded structure of four hairpins (peptide 1, GB1, TRPZIP2 and TRPZIP4): 1) the hydrophobic core cluster is well packed (the hydrophobic core radius of gyration R_g^{core} is less than 5.0 Å), 2) the all-residue C_α-RMSD from the corresponding NMR structure is less than 2.5 Å, 3) at least four backbone hydrogen bonds are formed ($N_{HB} \geq 4$, totally 6 hydrogen bonds for peptide 1, GB1 and TRPZIP4 and 5 for TRPZIP2 peptide). Here we define backbone hydrogen bonds HB 1-6 ordered from the terminal to the turn (for TRPZIP2, the hydrogen bond HB 1 (S1 O-K12 H) located at the terminal of the hairpin is not taken into account due to its high instability, therefore there are totally five hydrogen bonds involved in the data analysis which are from HB 2 to HB 6, see Fig. 3.1). A hydrogen bond is considered as formed only if the distance between the carbonyl oxygen and amide hydrogen (O...H) is less than 3.5 Å and the N-H-O angle is greater than 145^o.

The atomic-resolution structure is not available for either 20mer or its mutant 20mer^{D}P6D. Folded β-sheet structures herein are defined based on their 2D-NOESY (Nuclear Overhauser Enhancement Spectroscopy) spectra respectively.[85] The two 2D-NOESY spectra illuminated that both peptides form planar three-stranded β-sheets. The representative structures of the most populated folded states for two peptides were chosen from the simulation trajectories and shown in Fig. 3.2 (a & b), which display backbone hydrogen bond assemblies consistent with the analysis results of their NMR data. Hydrogen bond assemblies for hairpin 1 (N-terminal) and hairpin 2 (C-terminal) in both 20mer and 20mer^{D}P6D are organized in Table 3.1 and shown in Fig. 3.2.

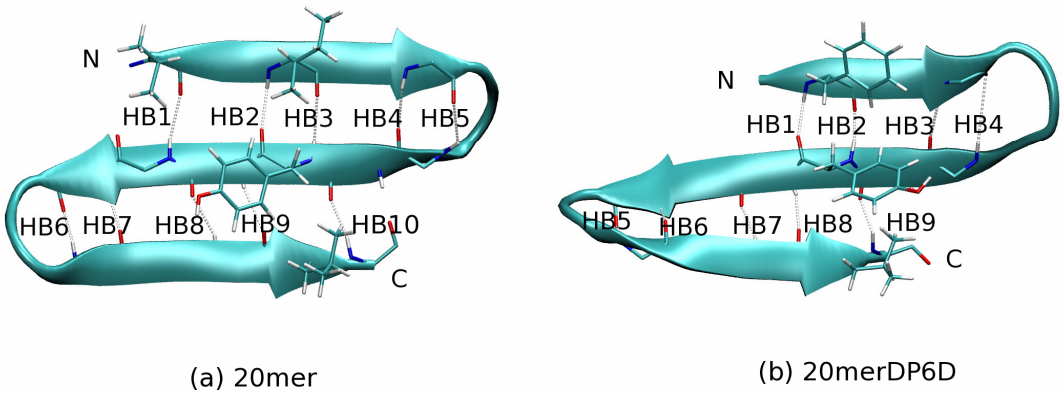


Figure 3.2. The representative structure of the folded structure of (a) 20mer and (b) 20mer^{D}P6D. Residues that form the hydrophobic core are shown in licorice mode (V1, I3, Y10 and L19 in 20mer, F2, Y10 and L19 in 20mer^{D}P6D) and backbone hydrogen bonds are shown in dash lines.

The folded structure of 20mer (20mer^{D}P6D) is defined herein with following criteria: 1) the hydrophobic core cluster is well packed (R_g^{core} <7.0 Å), 2) at least seven of the total backbone hydrogen bonds are formed (totally 10 hydrogen bonds for 20mer and 9 for 20mer^{D}P6D).

Table 3.1. Backbone hydrogen bond list for 20mer and 20mer^{D}P6D.

Peptide	Hairpin	Hydrogen bond	Peptide	Hairpin	Hydrogen bond
20mer	1	HB1: V1 O-E12 H HB2: I3 H-Y10 O HB3: I3 O-Y10 H HB4: S5 H-K8 O HB5: S5 O-K8 H	20mer^{D}P6D	1	HB1: F2 H-Y10 O HB2: F2 O-Y10 H HB3: T4 H - K8 O HB4: T4 O-K8 H
	2	HB6: V13 O-O16 H HB7: V13 H-O16 O HB8: T11 O-I18 H HB9: T11 H-I18 O HB10: T9 O-Q20 H		2	HB5: V13 O-O16 H HB6: V13 H-O16 O HB7: T11 O-I18 H HB8: T11 H-I18 O HB9: T9 O-Q20 H

3.3.2 The folding mechanism of β-hairpins

In the present study, we used the ITS simulation to study the folding of peptide 1, TRPZIP4, 20mer and 20mer^{D}P6D. For each peptide, several independent MD simulation trajectories were obtained, each starting from the totally extended structure and lasting several hundred nanoseconds. We obtained numerous folding and unfolding events in

the time scale of microseconds for each peptide (totally 57 folding events for peptide 1, 26 for TRPZIP4, 25 for 20mer and 20 for 20mer^{D}P6D). Some typical trajectories are shown in Fig. 3.3.

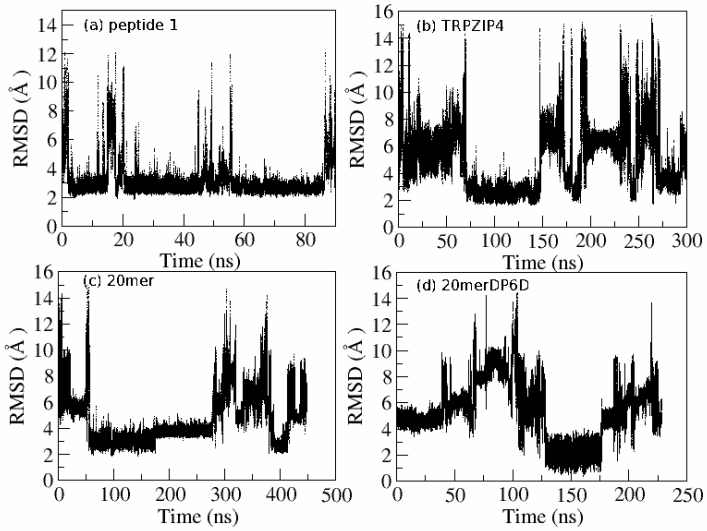


Figure 3.3. Time series of C_α-RMSD values in a typical trajectory from the ITS simulation for (a) peptide 1, (b) TRPZIP4 and (c) 20mer and (d) 20mer^{D}P6D.

To understand the folding mechanism of hairpins, we calculated the free energy landscape as a function of R_g^{core} and N_{HB}, as shown in Fig. 3.4, for peptide 1, TRPZIP2[161] and trpzip4. Previous studies, both experimentally and theoretically,[5,56,62,64,69-71,163] demonstrated that the folding of β-hairpin is generally a two-state process with the energy minima corresponding to the H state, in which the hydrophobic core is packed whereas no more than one hydrogen bond is formed, and the F (folded) state. This can be observed clearly from Fig. 3.4 and the free energy

landscape for GB1 (Fig. 3.2). Therefore starting from the extended structure, peptides quickly collapse to form a compact structure (H state). This process is largely barrierless and is expected to occur fast. Once peptides reach the H state, hydrogen bonds are formed by overcoming an energy barrier. Hydrogen bond formation is thus expected to be rate-limiting compared to the hydrophobic core collapse.

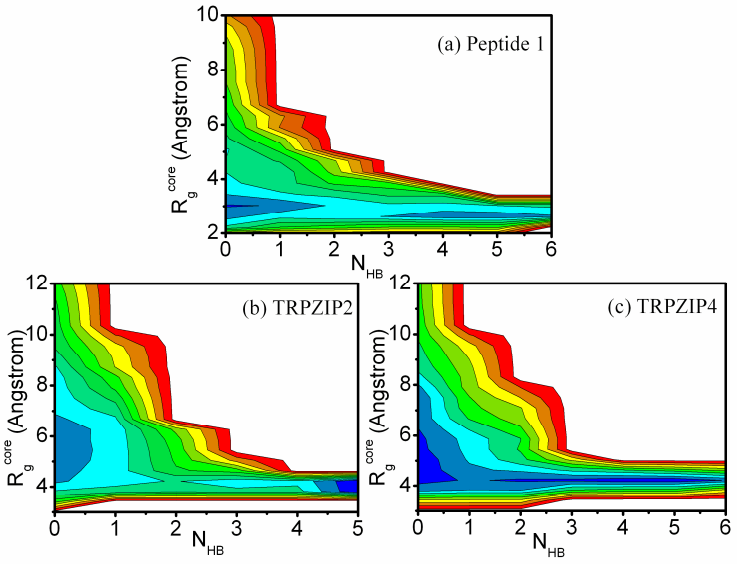


Figure 3.4. The free energy landscape as a function of R_g^{core} and N_{HB} for (a) peptide 1, (b) TRPZIP2 and (c) TRPZIP4. The contours are spaced at intervals of $2k_BT$.

One aim of the present study is to understand the effect of the amino acid sequence on the formation of turn structure and the hydrophobic core. We evaluate the formation of the turn configuration by the RMSD value of the turn segment compared to the corresponding fragment in the NMR structure ($RMSD_{turn}$): the turn structure is regarded as formed if $RMSD_{turn} \leq 1$ Å. We then calculated the free energy landscape as a function

of the number of the hydrophobic contacts (N_{HC}) and $RMSD_{turn}$ for all four related peptides and results are shown in Fig. 3.5.

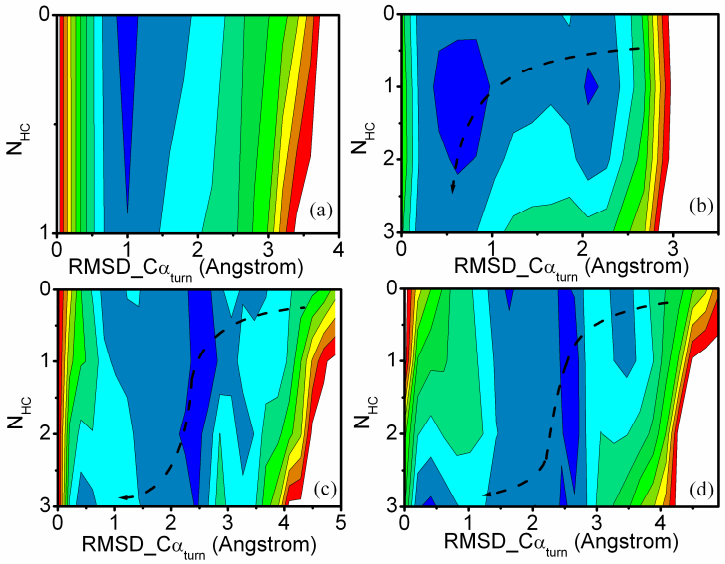


Figure 3.5. The free energy landscape as a function of N_{HC} and $RMSD_{turn}$ for (a) peptide 1, (b) TRPZIP2 and (c) GB1 and (d) TRPZIP4. The contours are spaced at intervals of $2k_BT$.

Peptide 1 may be regarded as a special case among the four peptides discussed here. It possesses only one pair of sidechain hydrophobic interaction in the native state (Ile5-Trp11) and a very stable β-turn as a result of ^{D}Proline. One can observe from the free energy profile of peptide 1 (Fig. 3.5 (a)) that there is only one local minimum ($RMSD_{turn}\approx 1$ Å and $N_{HC}\leq 1$) which shows that the turn formation is a spontaneous and barrierless process for peptide 1. Once the turn is formed it remains stable during the rest of the folding process whereas the sidechain interaction between two hydrophobic

residues is weak (N_{HC} swings between 0 and 1). Figure 3.5 (b) shows the free energy profile for the folding of TRPZIP2 which possesses a very stable type I´ turn structure and strong hydrophobic interactions among four tryptophans. In contrast to the free energy profile of peptide 1, two distinct local minima exist in this figure. One minimum ($RMSD_{turn}$≈2.2 Å and N_{HB}=1) represents the state at which only one pair of the hydrophobic core cluster is packed whereas the turn structure is not formed. The other broader minimum ($RMSD_{turn}$<1 Å and 0<N_{HB}<3) corresponds to the state at which both hydrophobic core and turn are formed. An energy barrier exists between the two local minima. Therefore for TRPZIP2 the packing of the hydrophobic core is largely facilitated by the formation of the β-turn. Without the formed β-turn only a portion of hydrophobic interactions can exist between the sidechains.

Free energy landscapes for GB1 and TRPZIP4 which have the disfavored type I β-turn (compared to type I´ turn) exhibit totally different features compared to those of peptide 1 and TRPZIP2. Besides the basin which represents the folded state ($RMSD_{turn}$<1 Å and N_{HC}=3), there is another distinct energy basin ($RMSD_{turn}$≈2.5 Å and 0≤N_{HC}≤3) in both free energy profiles. The latter and much broader basin corresponds to a state at which the hydrophobic core is well packed whereas the turn structure is not formed. This size difference between the two basins is opposite to that in the free energy profile of TRPZIP2. Therefore for GB1 and TRPZIP4, the hydrophobic core collapse is a barrierless process and an easy turn formation requires the packing of the hydrophobic core. The difference between peptide 1/TRPZIP2 and GB1/TRPZIP4 does reveal that the protein sequence influences the folding landscape and thus also the folding mechanism

of β-hairpins, although their final structures are similar. The sequence dependence is illustrated here by the competition between the turn formation and the hydrophobic core collapse. The introduction of a β-turn promoting sequence allows the easy formation of the turn. Otherwise the turn formation is associated with an appreciable free energy barrier and becomes lagged of the hydrophobic core formation (especially in the existence of strong hydrophobic interactions).

Next we calculated and analyzed the free energy landscape as a function of N_{HB} and $RMSD_{turn}$ to investigate the relation between the stability of backbone hydrogen bonds and the turn formation. The result is shown in Fig. 3.6. Free energy profiles of peptide 1 and TRPZIP2 possess similar patterns in which an energy barrier exists between two separated energy basins ($RMSD_{turn}\approx$ 1Å and $N_{HB}>3$, $RMSD_{turn}\approx$ 1Å and $N_{HB}<2$ for peptide 1; $RMSD_{turn}<0.8$ Å and $N_{HB}>2$, $RMSD_{turn}<2.5$ Å and $N_{HB}<1$ for TRPZIP2). This feature of the free energy profiles demonstrates that the turn structure formation is a barrierless process while the hydrogen bond formation has to overcome an energy barrier.

Free energy profiles of GB1 and TRPZIP4, however, again display different features. In both profiles, besides the basin representing the folded (F) state ($RMSD_{turn}<0.5$ Å and $N_{HB}\geq5$ for GB1, $RMSD_{turn}<0.5$ Å and $4\leq N_{HB}\leq6$ for TRPZIP4), there is anther energy basin (1.5 Å$<RMSD_{turn}<$3 Å and $N_{HB}\leq1$ for GB1, 1.5 Å$<RMSD_{turn}<$3 Å and $N_{HB}\leq2$ for TRPZIP4) which refers to the H state in the corresponding free energy landscape as a function of R_g^{core} and N_{HB} (see Fig. 3.4). There is also an additional basin existing in the profile of TRPZIP4 that represents a partially folded (P) state at which most hydrogen

bonds are formed without the turn formation ($RMSD_{turn}$≈2.5 Å and $4 \leq N_{HB} \leq 5$). The energy barrier between H and P states is obviously lower than that between H and F states. Therefore instead of folding directly from H to F state, it is more likely for GB1 and TRPZIP4 to form a stable intermediate, P state, from H state. Most of backbone hydrogen bonds are formed in this transition. The high energy barrier between P and F states demonstrates that the turn formation becomes the rate-limiting step during the folding process.

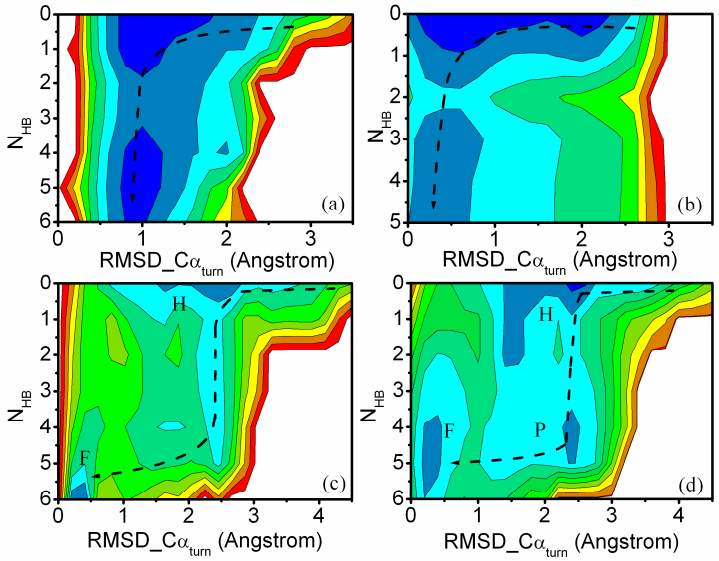


Figure 3.6. The free energy landscape as a function of N_{HB} and $RMSD_{turn}$ for (a) peptide 1, (b) TRPZIP2 and (c) GB1 and (d) TRPZIP4. The contours are spaced at intervals of $2k_BT$.

3.3.3 The stability of backbone hydrogen bonds

Through the result analysis of GB1 peptide, we observed that six backbone hydrogen bonds show different behaviors in the formation probability and stability.[164]

As shown in Fig. 3.7 (c), HB 6 (the hydrogen bond closest to the turn structure) manifests extreme instability during the folding process even when most hydrogen bonds are formed. The other five hydrogen bonds follow the order of 5>4>3>2>1 on the formation probability, corresponding to the hydrogen bond position from the inner to the outer of strands.

We then calculated the formation probability of backbone hydrogen bonds for other three hairpins and the result is shown in Fig. 3.7 (a, b and d). It is clearly seen from the figure that the hydrogen bond closest to the turn structure (HB 6) is the most possible to form when only one hydrogen bond is formed for peptide 1 and TRPZIP2. The easy formation of this hydrogen bond is consistent with the high stability of the turn structure in two peptides. However, with more hydrogen bonds formed, the formation probability of HB 6 drops and keeps low until the assembling of backbone hydrogen bonds is almost completed. The breaking of HB 6 when other hydrogen bonds are formed is a result of the constraint on the β-turn during the assembly of the native structure. On the contrary, HB 6 in GB1 and TRPZIP4 remains relatively unstable as revealed by its low formation probability in Fig. 3.7 (c & d). The instability of this hydrogen bond in GB1 and TRPZIP4 is a result of the instability of the β-turn structure. Therefore, one sees that the turn sequence has a direct effect on the relative stability of the nearby backbone hydrogen bond.

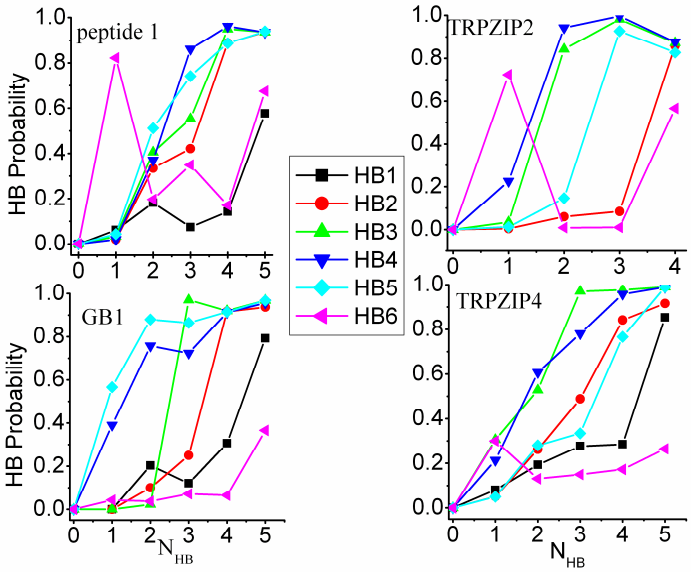


Figure 3.7. The probability for each hydrogen bond to be formed as a function of the total number of backbone hydrogen bonds formed for (a) peptide 1, (b) TRPZIP2, (c) GB1 and (d) TRPZIP4.

In addition to HB 6, Fig. 3.7 also shows different features on the relative stability of other four hydrogen bonds (HB 2 to HB 5) in different peptides. HB 1, the hydrogen bond near the terminal, however always shows very low stability during the assembling of backbone hydrogen bonds. As shown in Fig. 3.7 (a) and (c), the most stable hydrogen bonds in peptide 1 and GB1, which possess relatively few hydrophobic sidechains among four peptides, are HB5 and HB4. The former is close to the turn structure and the latter is surrounded by the hydrophobic core cluster. The difference in the two figures is that the stability difference among the four hydrogen bonds is smaller in peptide 1 compared to that in GB1 peptide. In contrast to peptide 1 and GB1, the stability of HB5 in TRPZIP2 and TRPZIP4, which possess the higher sidechain hydrophobicity, is

weakened and two hydrogen bonds (HB3 and HB4) in the middle of the hydrophobic core cluster become the strongest ones. Therefore the sidechain hydrophobicity also has an effect on the stability of the nearby hydrogen bonds.

3.3.4 The folding mechanism of β-sheets

To decrypt the folding mechanism of two β-sheets of 20mer and its mutant 20mer^{D}P6D, we first performed the hierarchical clustering analysis to characterize populated conformations obtained by MD simulations. The clustering analysis is based on C_α-RMSD (compared to the defined folded states as shown in Fig. 3.2) calculated between cluster structures and the cluster radius is selected as 6 Å. All simulation trajectories were involved with proper weighting in the clustering analysis. The most populated cluster structures are shown to represent the molten globule (H), the partially folded (P) and the folded (F) states.

To get a direct insight into the classification of different states, R_g^{core} and N_{HB} were calculated for each of the three states. On average, H state has a compact hydrophobic core cluster (R_g^{core} ≈6.0 Å) but with few native hydrogen bond formed ($N_{HB}\leq 1$), typical structures of H state are shown for 20mer and 20mer^{D}P6D in Fig. 3.8 (a & b) respectively. P state has a more compact global structure in which several native hydrogen bonds are formed ($3\leq N_{HB}\leq 5$ for both 20mer and 20mer^{D}P6D). F state is a state in which the hydrogen bond assembling is mostly complete ($N_{HB} \geq 8$ for 20mer or $N_{HB}\geq 7$ for 20mer^{D}P6D). It is worth mentioning that the clustering analysis provided several clusters of P state, in which the hairpin 2 (C-terminal) is always folded whereas hairpin 1 (N-terminal) is not.

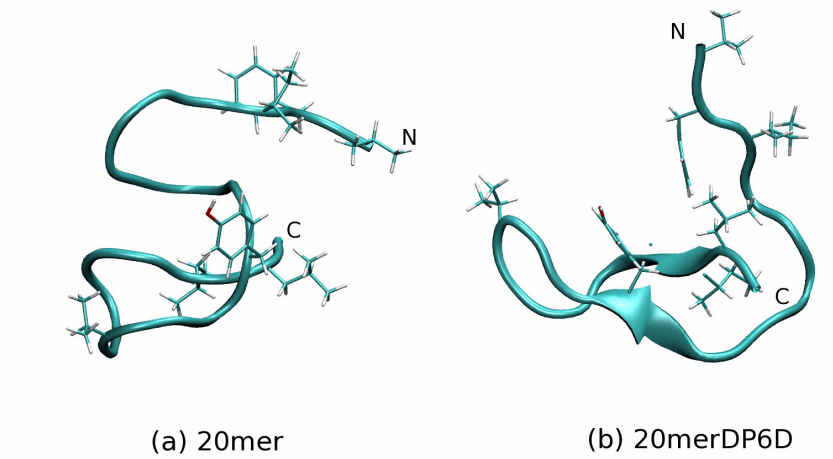


Figure 3.8. The representative structure of H state for (a) 20mer and (b) 20mer^{D}P6D from the clustering analysis. All hydrophobic residues are shown in licorice mode (V1, F2, I3, Y10, V13, I18, L19).

To thoroughly detect the conformations of the partially folded state, we chose the snapshots from all trajectories which satisfy the criteria of P state (R_g^{core}<7.0 Å and $3 \leq N_{HB} \leq 5$) and then run the hierarchical clustering analysis again on the limited ensemble for both 20mer and 20mer^{D}P6D. Three representative structures of P state are shown in Fig. 3.9 (a-c) (20mer) and Fig. 3.10 (a-c) (20mer^{D}P6D). Moreover, a minor cluster which represents a misfolded state is also achieved in the clustering analysis (4.86% in 20mer and 12.3% in 20mer^{D}P6D): hairpin 1 is folded with the proper hydrogen bond assemble whereas the second segment of the peptide forms a misfolded hairpin with one residue shifted in the direction towards the C-terminal, as shown in Fig. 3.9 (d) and Fig. 3.10 (d).

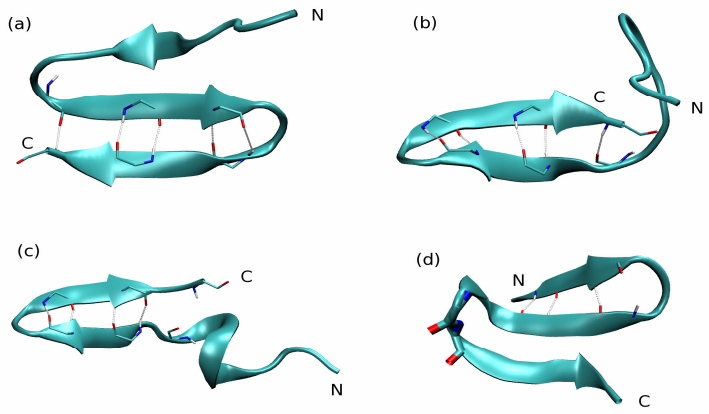


Figure 3.9. The representative structures of P state for 20mer. The shifted residues ^{D}P14 and G15 are shown in dynamic bond model and backbone hydrogen bonds are shown in dash lines.

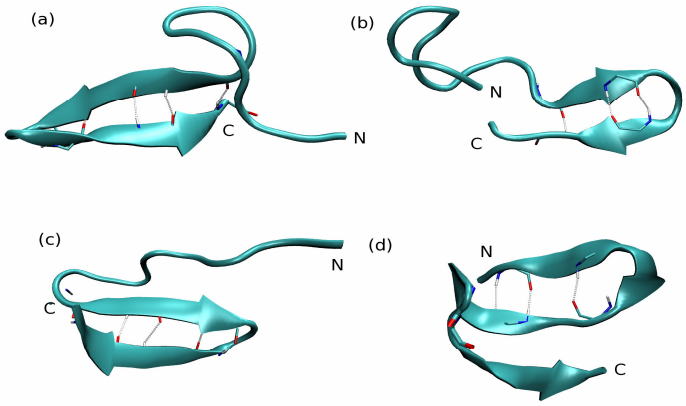


Figure 3.10. The representative structures of P state for 20mer^{D}P6D. The shifted residues ^{D}P14 and G15 are shown in dynamic bond model and backbone hydrogen bonds are shown in dash lines.

The free energy landscape as a function of R_g^{core} and N_{HB} was calculated for both 20mer and 20mer^{D}P6D, and is shown in Figs. 3.11 (a) and (b) respectively. These two free energy profiles have a similar shape, from which three distinct states (H, P and F states) are identified. Therefore the folding of two β-sheets is not a simple two-state process. Starting from the totally extended structure, each of the two peptides quickly folds into a compact structure (H state) driven by the hydrophobic core collapse. Once H state is formed, the peptide further folds into a stable intermediate (P state) driven by the intrinsic hairpin propensities of the residues in the strands (e.g., backbone hydrogen bond interactions). The folded structure of the peptide is finally achieved through the transition from P to F state, which has the highest energy barrier to overcome and therefore is the rate-limiting step.

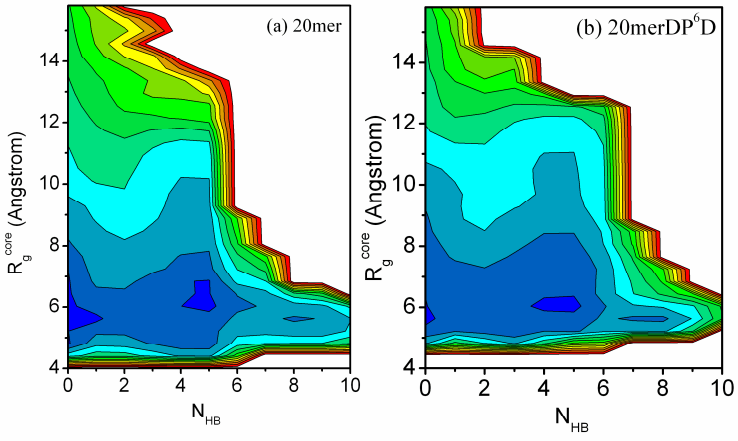


Figure 3.11. The free energy landscape as a function of R_g^{core} and N_{HB} for (a) 20mer and (b) 20mer^{D}P6D. The contours are spaced at intervals of $2k_BT$.

To further investigate the relation between the amino acid sequence and the stability of backbone hydrogen bonds we again calculated the formation probability of each backbone hydrogen bond as the function of the total number of formed hydrogen bonds and the result is shown in Fig. 3.12. Hydrogen bonds are named from the N-terminal to C-terminal. Therefore for 20mer HB 1-HB 5 belong to hairpin 1 (starting from the N-terminal) and HB 6-HB 10 belong to hairpin 2 (HB 10 being close to the C-terminal). For 20mer^{D}P6D of which the turn structure in hairpin 1 is altered there are totally four hydrogen bonds named HB 1- HB 4 in hairpin 1 and five hydrogen bonds named HB 5-HB 9 in hairpin 2.

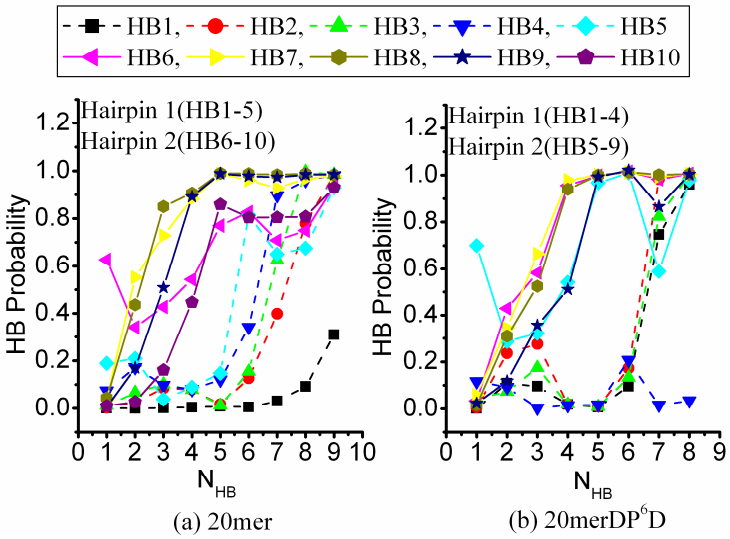


Figure 3.12. The probability for each hydrogen bond to be formed as a function of the total number of native backbone hydrogen bonds formed for (a) 20mer and (b) 20merDP^{6}D.

One can easily see the difference in the relative stability of hydrogen bonds belonging to different hairpins for each β-sheet: the formation probability and the stability of hydrogen bonds in hairpin 2 are apparently higher than those of hydrogen bonds in hairpin 1. This observation supports a three-state folding model for 20mer and its mutant. Moreover, it can be seen from Fig. 3.12 that in the hairpin with the strong β-turn promoting sequence the hydrogen bond closest to the turn (HB5 in hairpin1 and HB6 in hairpin 2 for 20mer, HB5 in hairpin 2 for 20mer^{D}P6D) is easy to form, as indicated by the high formation probability for corresponding hydrogen bonds when N_{HB} is equal to 1. On the contrary, the hydrogen bond closest to the turn in the hairpin 1 of 20mer^{D}P6D in the absence of ^{D}Proline (the strong turn promoter) does not have such a property. These results are consistent with the observation on the turn sequence dependence of the hydrogen bond stability in individual β-hairpins as described earlier. In addition, hydrogen bonds in the middle of strands have similar and high stability in both β-sheets (e.g., HB7, HB8 and HB9 in hairpin 2 of 20mer and HB6, HB7 and HB8 in hairpin 2 of 20mer^{D}P6D). Considering the fact that hairpin 2 in the folded structure of both β-sheets possesses the hydrophobic interaction only from the pair of Tyr10 and Leu19, this result demonstrates the same sidechain hydrophobicity dependence of the stability of backbone hydrogen bonds as that in peptide 1.

Besides backbone hydrogen bond interactions, the sidechain-sidechain interactions especially the hydrophobic interaction was shown to be another important factor which determines the hairpin stability.[78,80] Kyte and Doolittle scale is widely used to describe the hydrophobic character of amino acids,[165] in which more positive value of the

hydrophobicity index indicates higher hydrophobicity of the residue. In both β-sheets, hairpin 1 possesses four hydrophobic residues, Val1, Phe2, Ile3 and Tyr10, and the sum of hydrophobicity indexes is 10.2. Hairpin 2 possesses four hydrophobic residues, Tyr10, Val13, Ile18 and Leu19, with the corresponding hydrophobicity indexes added up to be 11.2. Therefore the sidechain hydrophobicity in hairpin 1 is lower than that in hairpin 2. As a result, in the well-defied folded structure for two β-sheets as shown in Fig. 3.2, sidechains on the middle strand (e.g., the aromatic group of the hydrophobic residue Tyr10) are biased apparently to the C-terminal and consequently the hydrophobic core cluster in hairpin 2 is more compact in comparison to that in hairpin 1. Therefore the hydrophobic interaction is again shown to be an important factor which controls the folding process of β-sheets and induces the folding priority of hairpin 2.

To further understand this influence of the sidechain hydrophobicity we calculated the free energy landscape as a function of Q_{H1} and Q_{H2} (Q_{H1} (Q_{H2}) is the fraction of hydrophobic contacts in hairpin 1 (hairpin2)) and the result is shown in Fig. 3.13. Two distinct local free energy minima exist in the profile of 20mer which refer to the partially folded state with only the hydrophobic core in hairpin 2 packed and the folded state respectively. The hydrophobic core collapse in hairpin 2 can be regarded as a barrierless process whereas that in hairpin 1 demands to overcome an energy barrier.

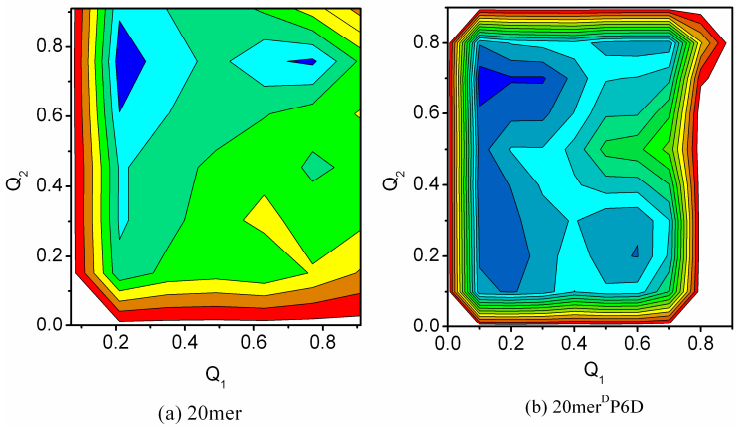


Figure 3.13. The free energy landscape as a function of Q_1 and Q_2 for (a) 20mer and (b) 20mer^{D}P6D. The contours are spaced at intervals of $1k_BT$.

There are four local minima in the free energy profile of 20mer^{D}P6D which refer to the unfolded state ($Q_1 \approx 0.2$ and $0.1 \leq Q_2 \leq 0.5$), two partially folded states ($0.1 \leq Q_1 \leq 0.3$ and $0.6 \leq Q_2 \leq 0.7$, $0.4 \leq Q_1 \leq 0.7$ and $0.1 \leq Q_2 \leq 0.3$) and the folded state ($0.5 \leq Q_1 \leq 0.7$ and $Q_2 \approx 0.8$) respectively. The hydrophobic core collapse in hairpin 2 is still a barrierless process. The energy barrier between the partially folded state with the compact structure of hairpin 2 to the folded state is obviously lower than that between the partially folded state with the compact structure of hairpin 1 to the folded state. Therefore for both 20mer and 20mer^{D}P6D, the hydrophobic core in hairpin 2 is easier to form compared to that in hairpin 1.

3.4 Conclusion

Through the analyses of the free energy landscape as the function of several types of reaction coordinates ($N_{HB,}$ N_{HC} and $RMSD_{turn}$) we observed that significant differences exist in folding free energy landscapes of four hairpins (peptide 1, GB1, TRPZIP2 and TRPZIP4). These differences in folding free energy landscapes are expected to result in different folding mechanisms of similar β-structures of varied sequences. Differences are essentially induced by the protein sequence and represented by the different relative stability of the hydrophobic core and the β-turn. In summary, the strong β-turn promoting sequence (peptide 1 and TRPZIP2) induces the easier formation of the turn structure compared to the hydrophobic core. The formation of the latter is largely facilitated by the former. Even in the presence of strong hydrophobic interactions, the hydrophobic core can be only partially packed without a formed β-turn structure. Accordingly, the hydrogen bond formation becomes the rate-limiting step. This is consistent with the modified "hydrogen-bond-centric mechanism" proposed by Tsai *et al.*[66] On the other hand, in the presence of the disfavored turn structure (GB1 and TRPZIP4), the hydrophobic core collapse becomes a barrierless process and thus easily occurs. Both backbone hydrogen bond formation and turn formation demand to overcome energy barriers. The energy barrier of the former is lower than that of the latter. Therefore the turn formation becomes the rate-liming step in the folding process. This is in agreement with the "hydrophobic-core-centric mechanism" of hairpin folding.

It is also observed in folding simulations of β-hairpins that both the turn structure and the sidechain hydrophobicity strongly affect the stability of their nearby backbone

hydrogen bonds. The hydrogen bond closest to the turn structure (HB6) exhibits the large idleness during most of the folding process. However, in the presence of strongly favored turn structure, this hydrogen bond could be formed easily (it is the most possible one to form when only one hydrogen bond is formed for the entire peptide although it becomes idle again when more hydrogen bonds are formed). This transient strong stability of the hydrogen bond closest to the turn might be related to the very stable turn structure which allows a close proximity of the residues forming this hydrogen bond. However when the hydrophobic core collapse occurs there is an essential requirement for the structural adjustment and as a result the hydrogen bond closest to the turn is broken quickly to allow the rearrangement of the rest of the β-structure. It is reformed only when most of other hydrogen bonds are formed and a native structure is achieved. On the other hand, in the presence of weak hydrophobic interactions, backbone hydrogen bonds except the two marginal ones (HB 1 and HB 6) have more or less even stability. The introduction of strong hydrophobic interactions result in the stronger stability of the two hydrogen bonds in the middle of the hydrophobic core cluster (HB 3 and HB 4) compared to other ones. Therefore it is speculated that strong hydrophobic interactions provide a more hydrophobic environment and thus strengthen nearby hydrogen bonds. In summary, the order of the formation probability and the stability of backbone hydrogen bonds except HB 6 are: 4,5>3>2>1 (peptide 1), 3,4>2,5>1 (TRPZIP4) and 4>3>5>2 (TRPZIP2).

The similar turn influence on the stability of its nearby hydrogen bond is also found in two β-sheets, 20mer and the mutant 20mer^{D}P6D. In simulations, both peptides form

steady folded states with proper hydrogen bond assemblies based on their 2D-NOESY specta.[85,86] The mutation of D-Proline in hairpin 1 of 20mer dramatically changes the structure of the corresponding segment, which folds into a stable hairpin with a type I turn composed by five residues rather than the 2:2 type II´ turn promoted by D-Proline residue. Through MD simulation studies of 20mer and 20mer^{D}P6D, we observed that the folding of a three-stranded β-sheet is a three-state transition corresponding to the hydrophobic core collapse, the folding of hairpin 2 (C-terminal) and the folding of hairpin 1 (N-terminal). The hydrophobic core collapse is a barrierless process. The formation of hairpin2 is easier than that of hairpin 1 which be explained by the higher sidechain hydrophobicity in the former, again providing evidence for the effect of sidechain hydrophobicity on the strength of backbone hydrogen bonds.

CHAPTER IV

THE CHEMOMECHANICAL COUPLING MECHANISM OF KINESIN*

4.1 Introduction

Conventional kinesin walks by a hand-over-hand mechanism.[112,116] Using high resolution single molecule fluorescence measurement, it was found that during walking the displacement of each of the two heads of the homodimeric kinesin alters between ~16 nm and 0 nm, and the two heads exchange leading and trailing positions with each 8-nm stepping of the kinesin center for each ATP hydrolysis.[112] The hand-over-hand mechanism requires highly cooperative motions of kinesin in which the two identical motor heads alternate their roles in catalytic reactions and in physical motions. In principle, the hand-over-hand mechanism of kinesin can be either symmetric or asymmetric. Due to the lack of a mirror symmetry and the fact that the central stalk is formed from a coiled-coil, in symmetric walking each of the forward step should accompany a 180° rotation of the central stalk. However, in the asymmetric walking mechanism, no rotation of the stalk region should occur and as a result the steps differ depending on which head takes the leading position. The experiment of Hua *et al*. showed that kinesin walks in the absence of the rotation of the central stalk.[114] Combined with the later results that kinesin walks by a hand-over-hand mechanism, this experimental observation indicates that kinesin walks by an asymmetric mechanism.

---

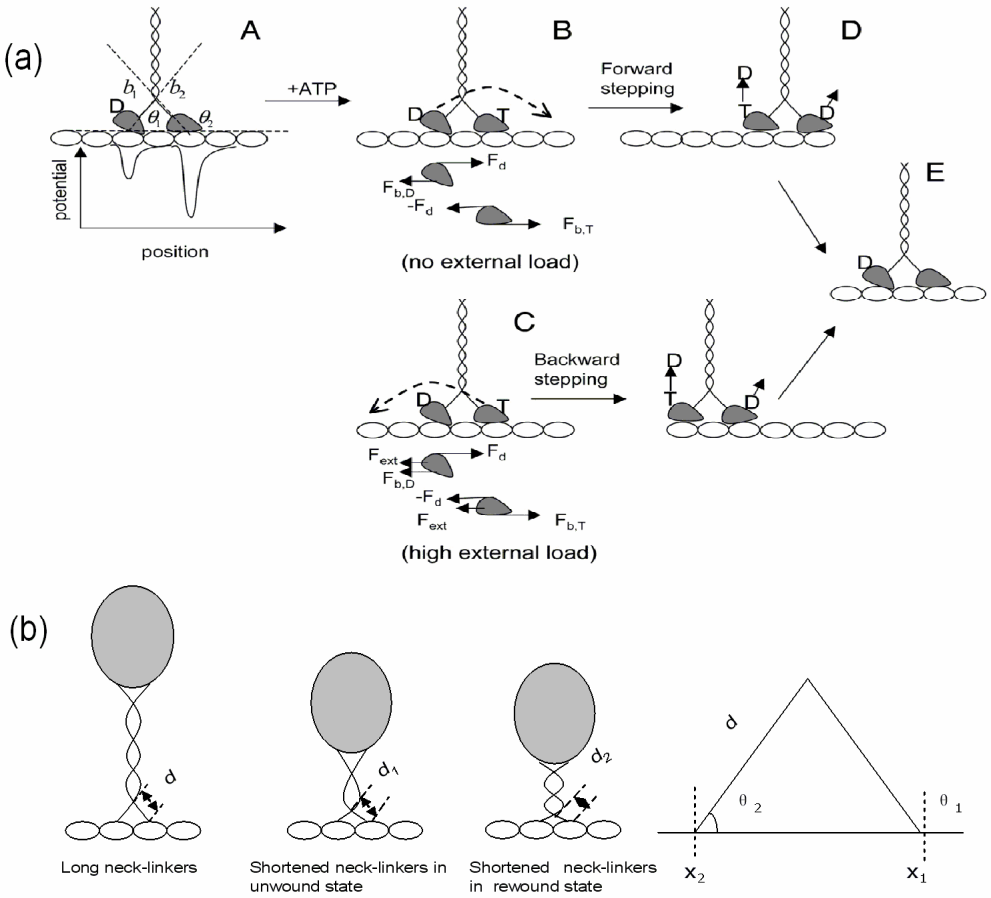
* Parts of this chapter are reproduced with permission from "On the Hand-over-hand Mechanism of Kinesin" by Shao Q., Gao Y. Q., *Proc. Natl. Acad. Sci. USA* 2006, 103, 8072-8077, Copyright 2006 National Academy of Science, U.S.A and from "Asymmetry in Kinesin Walking" by Shao Q., Gao Y. Q., *Biochemistry* 2007, 46, 9098-9106, Copyright 2007 American Chemical Society.

The asymmetry in the hand-over-hand mechanism is not manifested in the processive motion of the conventional kinesin, but revealed by the "limping" phenomenon during the walking of the homodimeric kinesin with the truncated coiled-coil stalk region (the stepping rates change for each alternating step).[118] The asymmetry property was also confirmed by other experimental observations, e.g., although the depletion of the ATP hydrolysis activity of one of the kinesin heads leads to the large decrease of the speed of the kinesin mutant, it can still walk along the microtubule[120]. In the present study, we proposed a physical model to investigate the asymmetric hand-over-hand bidirectional movement of kinesin.[166] In particular, the force dependence of the forward/backward ratios, of dwell times, and of the coupling between ATP hydrolysis and kinesin walking was studied. The asymmetry during kinesin walking is included in the model by assuming that lengths of neck-linkers are dependent on which kinesin head takes the leading position.[167] This difference in lengths of neck-linkers is trivial and does not induce asymmetry during the walking of the conventional kinesin (with relative long neck-linkers) whereas is vital for the kinesin mutants and therefore leads to the limping. We also tried to model the stepping of the heterodimeric kinesin mutant with the ATP hydrolysis activity in one head depleted and the stepping of the wild-type kinesin in the presence of both ATP and AMPPNP.

4.2 Theoretical Model

The proposed model of kinesin consists of two heads that are connected through their neck-linkers (see Scheme 4.1 (a)).

Scheme 4.1



In Scheme 4.1, (a) presents a schematic model for kinesin and its forward/backward stepping mechanism.[166] State A shows kinesin at the pre-stroke state. ATP binding (state A to state B and state A to state C) generates a power stroke that exerts a force (F_d) to the plus end for the rear head and a force ($-F_d$) to the minus end for the front head. When the external force is small (state B), because the front head binds the microtubule more strongly, the front head keeps its position whereas the rear head moves forward

(state B to state D). Because the external force is in the direction opposite to the power stroke on the rear head, these two forces (also the binding force at the microtubule site, $F_{b,D}$) tend to cancel each other, and at large external forces (state C), the rear head holds its position. The external force and the power stroke are both in the minus direction for the front head, and, together, they pull this head off its binding site (against the binding force $F_{b,T}$) and move it in the minus direction (state C to state E). State E is the same chemical state as state A but is generated after either a forward or backward stepping, and ATP binding to state E will lead to either a forward or backward stepping, depending on the external force.

In Scheme 4.1(b) left, three schematic diagrams show the effects of the shortened neck-linkers. From the left to right the three diagrams represent a wild-type kinesin, a kinesin with a shortened neck-linker due to the truncation of the coiled-coil and a kinesin with the further shortened neck-linkers due to rewinding of the coiled-coil. The graphic in Scheme 4.1 (b) right shows the triangle formed between two neck-linkers and the microtubule. The angle formed between neck-linker and microtubule is a function of leg length d and the relative head position x_1-x_2: $cos\theta_1=(x_2-x_1)/d/2$, $\theta_2=\pi-\theta_1$.

Each head binds to the microtubule with an affinity that depends on its chemical state. The chemical state determines the conformation between the head domain and its neck-linker domain, and therefore is coupled to the binding and motion of the other head. Kinesin at different chemical states binds the microtubule with different affinities: the binding affinity is the largest for the post-hydrolytic ADP/Pi state; the ATP state and the empty state, based on detachment experiments,[102,103] are assumed to have similar but

smaller binding affinities; the ADP state of kinesin head binds the microtubule the most weakly. The main assumption of the present model is that the ATP (or AMPPNP) binding induces a conformational change of the front kinesin head and thus provides a driving force in the plus direction[94]. It is further assumed that the hydrolysis of ATP to ADP/Pi generates an additional driving force for the motion of kinesin, even though this force could be significantly smaller than that due to ATP binding and the ATP and ADP/Pi bound states take similar stable conformations. A conformational change due to the hydrolysis of ATP has been suggested by fluorescence polarization experiment[100], which showed that the post-hydrolysis state (mimicked by ADP.AlF_4^-) tends to be more rigid than the pre-hydrolysis state (AMPPNP bound), in particular for a monomeric kinesin[168]. The pre-hydrolytic state, in accordance with experiments,[100,168] is assumed to bind microtubule more weakly than the post-hydrolytic state. This assumption doesn't affect the results on the wild-type kinesin, the hydrolysis of ATP by which is fast enough for the population of the pre-hydrolytic state to be neglected. But it does have an influence when the hydrolysis of ATP is missing, either due to the lack of the ATPase activity of the catalytic core or due to the substitution of ATP by AMPPNP, in which the driving force due to ATP hydrolysis is absent. To take into account the two-step force generating mechanism by ATP binding and hydrolysis, four possible chemical states are included for each kinesin head (ATP, ADP/Pi, ADP, and empty). For the conventional kinesin, since ATP hydrolysis occurs fast, only three states were considered. The ATP state is taken into account explicitly in the present model due to the lack of the

hydrolysis of AMPPNP, or ATP by a mutant motor domain. The pre-hydrolytic ATP state is assumed to be the same as the AMPPNP state.

The prevailing speculation on the causation of the asymmetry in kinesin walking is that the central coiled-coil stalk formed by the two kinesin monomers changes conformation during the stepping of the kinesin, so that the lower portion of the coiled-coil goes through unwinding and rewinding cycles, depending on which head is taking the leading position.[92,167] As a result, the length of neck-linkers varies with the exchange of the head positions, although the neck-linkers from the two kinesin heads always have equal lengths (so that the lengths of the two neck-linkers change simultaneously during kinesin walking). This change in the neck-linker length is included in our simplified model by the change of the "leg length" parameter d (see Scheme 4.1 (b), the total length from the lower end of the coiled-coil to the head/microtubule binding site), which is 5.2 nm in the unwound and 4.7 nm in the rewound state of the coiled-coil of a wild-type (conventional) kinesin[167]; the changes in the lengths of the two neck-linkers sum up to ~ 1 nm, the approximate length of one repeat of the coiled-coil.[118] As we mentioned above, this length change (between 5.2 nm and 4.7 nm) has little influence on the walking of the wild-type kinesin, and there is no obvious "limping" at all tested external loads.

We construct a similar model for a truncated kinesin, of which the upper portion of the coiled-coil stalk is shortened. This shortened central stalk is speculated to make the lower end of the stalk to wrap further (see a more detailed explanation in Scheme 4.1 (b)), and consequently, the neck-linkers become shorter.[118] This length change in neck-linkers is also expected to be dependent on how long the stalk is shortened: the more the

stalk is shortened, the shorter are the neck-linkers.[118] As suggested by Block and coworkers,[118] the change in the neck-linker length during the walking of kinesin could also be a result of the switch between the registered and misregistered coiled-coil near the neck region. In the present study, due to the lack of detailed structural information, we do not distinguish between the possible origins for the change of the length of the neck-linkers during kinesin walking. In the model for the truncated kinesin the value of d (for both unwound and rewound states) is smaller than that for the wild-type kinesin. We note again that although the neck-linkers are assumed to take different lengths depending on which head takes the leading position, the neck-linker lengths of the two motor domains are always of equal lengths, changing simultaneously during kinesin stepping. The asymmetry of kinesin walking discussed here refers to the difference between any two consecutive steps.

Once the model is constructed, simple potential energy functions are used to describe the interaction between each of kinesin heads and the microtubule binding sites as well as that between the two kinesin heads.

4.2.1 Potential energy functions

The two kinesin heads in the present model are represented by two points with their position given along the microtubule, which is represented by a single line. The two neck-linkers, with always equal lengths, are connected through a connecting point which is determined by the length of the neck-linkers (the distance between the connecting point B and the head position A and A′ (see Scheme 4.1 (b)). Each of the two heads can bind the microtubule, with the binding sites separated by 8.1 nm on the line representing

the microtubule. The binding between the kinesin head and the microtubule is represented by a simple function form

$$V_b = V_{b,0}^s \left(1 - \sum_i \exp(-(\frac{x - r_i}{a})^2) \right) \tag{4.1}$$

In Eq. 4.1, $V_{b,0}^s$ and α are parameters define the depth and width of the binding potentials. The width of the binding potential is assumed to be independent of the chemical state of kinesin (defined by its occupation: empty, or occupied with ADP, ADP/Pi, or ATP).

The depth of the potential, thus the binding affinity, is dependent on the chemical state of the kinesin head, the $V_{b,0}^s$ used in the present study is 4, 7, 7, and 9 kcal/mol for the ADP, empty, AMPPNP (or pre-hydrolytic ATP), and ADP/Pi states, respectively (see Table 4.1). Besides binding affinities, to take into account the conformation preference of kinesin at each chemical state listed above, the two heads are assumed to interact through the neck-linkers, which are not necessarily rigid. The interaction between the two kinesin heads are again determined by chemical states: the stable conformation between each head and its linkage is assumed in the present model to be determined only by the chemical state of this head itself, thus each head independently generates a potential as a function of the relative positions of the two heads, $x_1 - x_2$.

Table 4.1. Parameter values used in the model.

Neck-length d_1 / d_2 (nm)*	Conventional kinesin Kinesin with shorter neck-linkers	5.2/4.7 (unwinding/rewinding state) 4.6/4.1 (unwinding/rewinding state)
Kinesin-microtubule binding affinity (kcal/mol)	ATP-bound $V_{s,0}^{ATP}$ Empty $V_{s,0}^{E}$ ADP/P1.bound $V_{s,0}^{ADP/Pi}$ ADP-bound $V_{s,0}^{ADP}$	7.0 7.0 9.0 4.0
Chemical transition Rate constant	ADP release (s^{-1}) ATP binding ($\mu M^{-1}s^{-1}$) ATP hydrolysis (s^{-1}) ATP release (s^{-1})	260/2.6 (front/rear head) 3.0/0.3 (front/rear head) 8.0/800.0 (front/rear head) 10.0
The width of the binding potential α (nm) (Eq.4.1)		0.45/0.20 (forward/backward)
Parameters in Eq. 2	Internal force constant K_b (k_BT / nm^2) $\Delta x_1(\Delta x_2)$ (nm)	0.047/0.04/0.031 (ADP·Pi/ATP/Empty state) -8.1 (ADP/Pi & ATP state) 8.1 (Empty state)
Overstretching potential V_0 (Eq.4.3)		10 k_BT
Chemical transition constraint	Chemically allowed range within binding site (nm) Optimal neck-linker angle Chemically allowed angle deviation for ATP binding	0.16 40º ±5 º
Diffusion constant D (nm^2/s)		2.0×10^4

*The exact value of the stalk length is not used in the present study. However, the stalk length change is inverse to the neck-linker length change during the kinesin walking: the longer the neck-linker becomes, the shorter the stalk is.

Therefore, this interacting potential consists of two terms, one from each head. For simplicity, this interacting potential is taken to be a quadratic form and the stiffness of the potential is assumed to be proportional to the binding affinity between this kinesin head and the microtubule. In the ADP occupied state, it is assumed that there is no orientation preference of the neck-linkers and thus this energy term is absent for the ADP state. In contrast, the ADP/Pi or AMPPNP head prefers to be the rear head, so if this head has a position of x_1, a quadratic potential in the form of $\frac{1}{2}K_s(x_2 - x_1 - 8.1)^2$ is added to the total potential, with the force constant K_s given below.

The total potential due to the conformational preference of the two heads are then

$$V_p = \frac{1}{2}K_{s,1}(x_1 - x_2 - \Delta x_1)^2 + \frac{1}{2}K_{s,2}(x_1 - x_2 - \Delta x_2)^2 \tag{4.2}$$

K_s is taken to be $0.047 k_B T / \text{nm}^2$, $0.04 k_B T / \text{nm}^2$ and $0.031 k_B T / \text{nm}^2$, for ADP/Pi, ATP, and empty states, respectively. Δx_1 and Δx_2 are determined by the chemical state of the kinesin head. For example, when head 1 is empty, $\Delta x_1 = 8.1\,\text{nm}$, and when it is occupied by ADP/Pi or ATP (AMPPNP) it is −8.1 nm.

In addition to the binding potential and the head conformation potential mentioned above, to include the excluded volume effect of the two heads and to avoid overstretching of the distance between the two heads, a potential of the form

$$V_{rep} = V_0\left[\exp\left((4 - |x_1 - x_2|)^2\right) + \exp\left((|x_1 - x_2| - 9)^2\right)\right] \tag{4.3}$$

is also included. V_0 is taken to be 10 k_BT.

Once the potential (the sum of the potentials in Eqs. 4.1, 4.2 and 4.3) is specified, the force on each head is taken as the negative derivative of the potential over the position of the head (e.g., $-\partial V / \partial x_1$, for head 1) and the propagation for the motion of the head is performed using the Euler form of the Langevin Equation:

$$x_1(t+\Delta t) - x_1(t) = (F_{x_1,i} / \zeta_{x_1,i})\Delta t + \Delta x_1(t), \tag{4.4}$$

and

$$x_2(t+\Delta t) - x_2(t) = (F_{x_2,i} / \zeta_{x_2,i})\Delta t + \Delta x_2(t). \tag{4.5}$$

$\zeta_{x_1,i}$ and $\zeta_{x_2,i}$ are the friction constants for the motion along x directions. The state i is defined by the occupation of the two heads. The terms $\Delta x_1(t)$ and $\Delta x_2(t)$ are the random displacement due to the stochastic force and in the presence of white noise, they can be written as $\Delta x_1(t) = \sqrt{2D_{x_1,i}\Delta t}R_1$ and $\Delta x_2(t) = \sqrt{2D_{x_2,i}\Delta t}R_2$, respectively. $D = \zeta / k_B T$, k_B being the Boltzmann constant and T being the temperature, is the diffusion coefficient and R's are random numbers. R_1 and R_2 are independent and each has a Gaussian distribution with a zero mean and a unitary standard deviation. The temperature is taken to be 298K and a value of 2×10^4 nm^2s^{-1} is used for the diffusion constant.

4.2.2 Chemical transition rate constants and their force dependence

In addition to the physical motion, chemical transitions occur between the different chemical states of each head (e.g., *ATP→ADP/Pi*, *ADP/Pi→ADP et al.*). Transitions between chemical states are treated as random processes and the probabilities for the transition occurrence are determined by rate constants of transitions.

The external force applied on kinesin inevitably results in the strain in the kinesin, which has an influence on the various chemical transitions.[90] The force dependence of the rate constants of chemical transitions is written in the present model as:

$$k = k_0 \exp(-|F - F_0|\delta / k_B T) \tag{4.6}$$

$\delta / k_B T$ is taken to be about 2.5 pN^{-1} and F_0 is taken to be about 5 pN to obtain a good fit to the experimental data on the force dependence of the kinesin speed, which shows a maximum at an external load F_{ext} =5 pN. Since the force-dependence of the binding affinities is not available, for simplicity, force-independent equilibrium constants are assumed in the present study. Therefore, Eq. (4.6) is used for all chemical transition processes. A more elaborate model should take into account the difference in the force dependence of different rate processes. The rate constant k_0 of the various chemical transitions are taken from biochemical studies and then slightly modified in the present research to best-fit the experimental data. In order to take into account the preference of kinesin walking towards the plus-end of microtubule, it is further assumed that the rate constants of chemical transitions are not uniform on the two heads. Chemical transitions, e.g., ATP binding and ADP release are much faster on the front head whereas the ATP hydrolysis occurs fast on the rear head instead of front head. Rate constants used in the present study are listed in Table 4.1.

In addition to the explicit force dependence, chemical transitions are also affected by the conformation of kinesin heads, represented here by the angle between the neck-linker and microtubule and the distance between the kinesin head and its binding site on microtubule. Chemical transitions are only allowed when the corresponding kinesin head

is within a certain distance (taken as the value of 0.16 nm in the present paper) from a binding site and when the neck-linker angle is close (±5º) to the stable value of the reacting site. When the angle is more than ±5º away or the distance is greater than 0.16 nm from its stable conformation, rate constant is decreased to 100 time smaller, the exact values of which have a small influence on the final results. Since in this model the force and kinesin conformation dependence is the same for all reactions (forward and backward), the equilibrium constants are independent of force or the kinesin docking or binding mode (a more elaborate model should certainly take into account the variance of the equilibrium constants).

4.3 Results

4.3.1 On the hand-over-hand walking of the conventional kinesin

First, we studied the translocation of the conventional kinesin along the microtubule in the absence of external forces. The relevant parameters used in the model are shown in Table 4.1. The neck-linker length for is 5.2 nm in the unwound and 4.7 nm in the rewound state of the coiled-coil for the conventional kinesin. Kinesin starts with one head in the nucleotide-free state and the other occupied with ADP. Trajectories of both kinesin heads obtained in such a condition and under different external forces (in both forward and backward directions) are shown in Fig. 4.1.

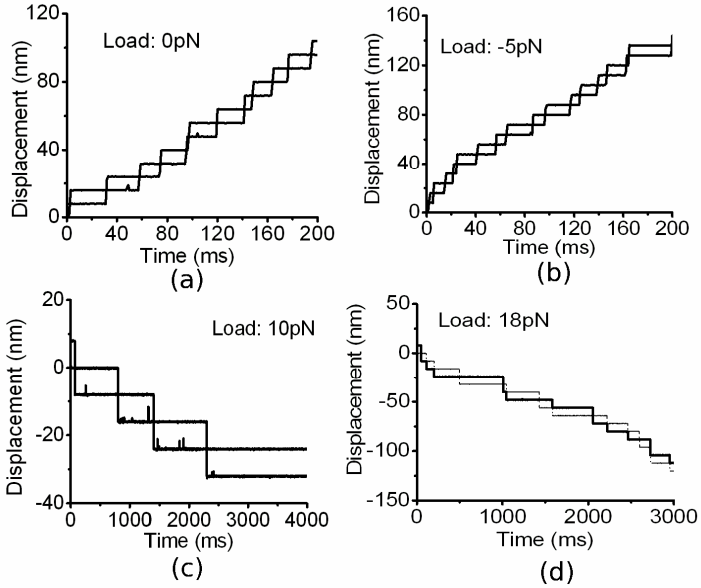


Figure 4.1. Trajectories of the conventional kinesin in the absence of external force (a) and in the presence of -5 (b), 10 (c) and 18 (d) pN. Trajectories of the two kinesin heads are shown. The ATP concentration is 1 mM.

One can see from Fig. 4.1 (a) that in the absence of external load, kinesin takes consecutive forward steps (to the plus end of the microtubule), and no backward steps are observed. It can be seen that during the walking of kinesin, the two heads take turns being the leading head. Therefore, kinesin walks by a hand-over-hand mechanism. The maximum speed of kinesin obtained at saturating concentrations of ATP is ≈400 nm·s^{-1}, and the Michaelis–Menten constant (K_M), at which ATP concentration the speed of kinesin is one-half of its maximum value, is ≈25 μM. Both parameters are in reasonable agreement with the experimental data of Nishiyama *et al.*[115] but are approximately two and four times, respectively, too small compared with the results of Block and

coworkers.[111] This difference is likely caused by the different diffusion constants of kinesin in the experimental setup or the different ATP-binding and release constants. The exact origin of difference is still under investigation. Parameters used in this study allow us to obtain a zero-force speed that falls on the experimental curve of Ref. 116.

When a force is applied in the minus direction and when it increases, kinesin takes more frequent backward steps (Fig. 4.1 (b); -5 pN). In the presence of a large enough force, one observes (Fig. 4.1 (c); +10 pN) sustained backward motions of kinesin with an 8-nm step size and with the two heads alternating the leading position. Therefore, in the presence of large forces, the calculations show a hand-over-hand walking of kinesin to the minus end of the microtubule. Although the speed of the backward motion increases with the force, the motion is slow compared with the forward steps in the absence of external load (≈30 nm/s at +15 pN compared with ≈400 nm/s at 0 pN).

4.3.1.1 Force and ATP concentration dependence of the kinesin speed

To further characterize the external force dependence of kinesin walking, the speed of kinesin is calculated as a function of the external force with 1 mM or 10 μM ATP in solution. The results are shown in Fig. 4.2. As shown in Fig. 4.2 (a), the applied force influences the kinesin motion in a rather complicated way.

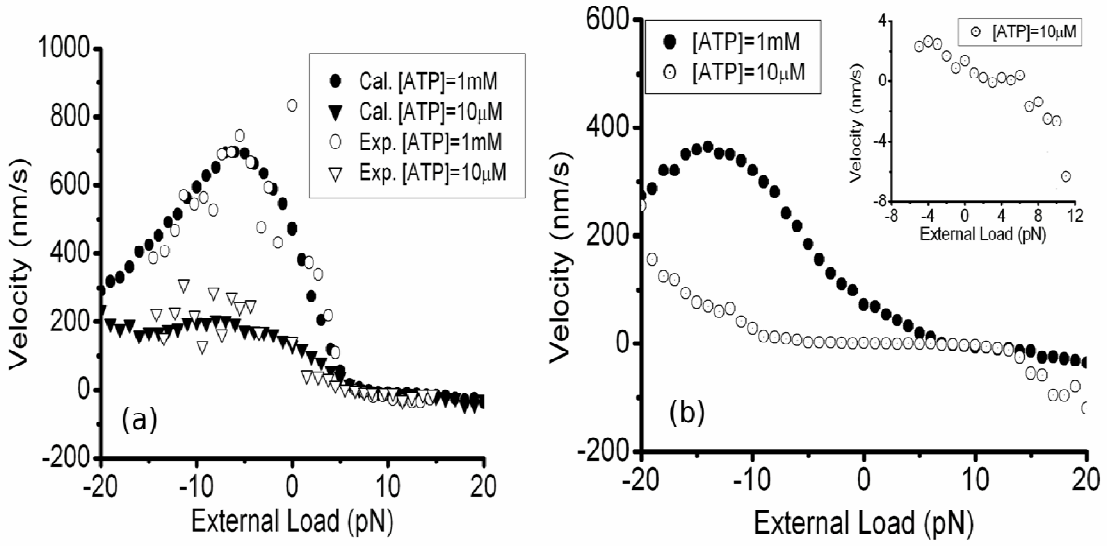


Figure 4.2. The velocity of kinesin as a function of the external force for (a) the conventional kinesin and (b) truncated kinesin (kinesin with the shorter neck-linker).

Although in general the increase of the opposing force decreases the kinesin speed and induces motions in the minus direction, the assisting force does not always increase the kinesin speed. Instead, at an ATP concentration of 1 mM, the speed of kinesin reaches a maximum in the presence of a −5 pN external load. Further increase (more negative) of the assisting external load decreases the kinesin walking speed (see Fig. 4.2 (a)). This observation is consistent with the experimental results of Carter and Cross.[116] In contrast, the speed of kinesin decreases monotonically when the external force changes from 0 to ≈7.5 pN (for [ATP] = 1 mM), without changing the direction of net motion.

4.3.1.2 Forward/backward step ratio and its force dependence

To understand the transition between forward and backward motions, we studied the external load dependence of the ratio between the numbers of forward and backward steps. Calculations were performed for both 1 mM and 10 μM ATP.

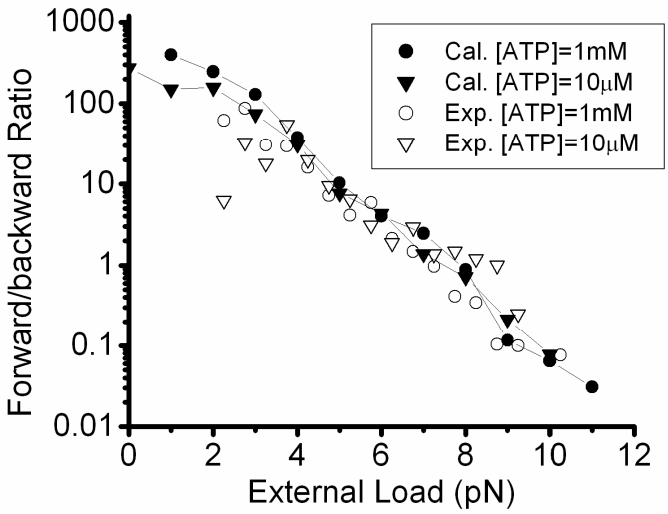


Figure 4.3. The forward/backward step ratio as a function of external force. The ATP concentrations are 1mM and 10μM.

As seen from Fig. 4.3, the difference due to the change of ATP concentration is small, and, in both cases, the forward/backward ratio decreases monotonically with the external force. This ratio becomes unity at the stall force (the force at which the speed of kinesin vanishes). Therefore, the zero speed of kinesin results from equal numbers of forward and backward steps without cease of motion. The stall force appears to be insensitive to the ATP concentration. The stall force decreases from ≈7.5 pN at 1 mM

ATP to 7 pN at 10 μM ATP (data is not shown here). These results are in reasonable agreement with the experiments. Both calculated and experimental results show an almost linear dependence of the forward/backward ratio as a function of the external force.

4.3.1.3 Dwell times and their dependence on the external force

Calculations were also performed to study the force dependence of the dwell time between individual steps for both forward and backward motions. This calculation is possible because the physical motion of kinesin heads occurs in a time scale of 10 μs, much faster than chemical transitions[116] (see trajectories in Fig. 4.1). Calculations were again performed for ATP concentrations of 1 mM and 10 μM. The calculated results are shown in Fig. 4.4.

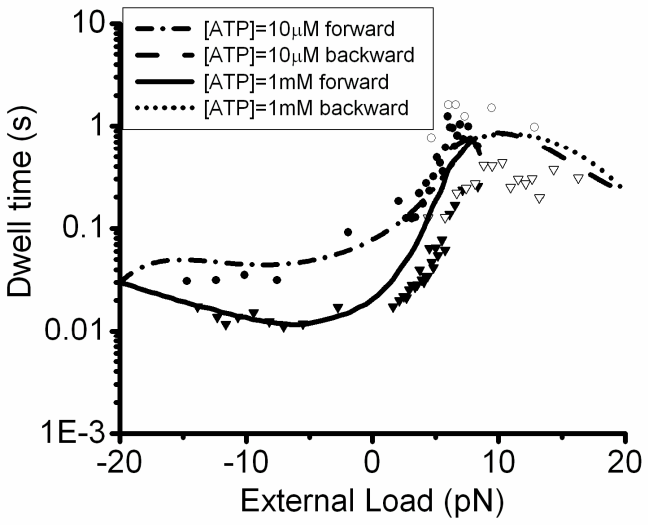


Figure 4.4. The dwell times for the backward and forward steps as a function of the external force. The ATP concentrations are 1mM and 10μM.

In general, the dwell time is smaller for 1 mM ATP than for 10 μM ATP, indicating that, at least at 10 μM, ATP binding is still rate limiting, consistent with $K_m \approx 30$ μM. In the force range of −15 to 5 pN, the dwell time is insensitive to the external force, although it appears that a minimum of dwell time exists at ≈−5 pN, consistent with Fig. 4.2 (a), which shows that maximum speed occurs at ≈−5 pN. Because of the rareness of backward steps, dwell time of backward steps was calculated only for forces >2 pN. When the external force increases from 5 to 10 pN, we see a sharp increase of the dwell time from ≈0.05 to ≈1 s for [ATP] = 1 mM and from ≈0.1 to ≈1 s for [ATP] = 10 μM. There is no apparent difference in the force dependence of forward and backward motions, indicating that forward and backward steps share same rate-limiting steps. The calculated force dependence of dwell times agrees reasonably well with experiments. However, unlike the experimental results, the calculated dwell times at large hindering forces (>10 pN) are independent of ATP concentrations in the range of 10 μM to 1 mM. The discrepancy between the calculated and experimental results is likely due to an additional force term that is missing in the present model. This ATP concentration dependence of dwell time indicates that K_M is much higher than 10 μM in this force range, even when the turnover rate of ATP becomes very small because of the slow translocation speed of kinesin. A K_M increasing with the hindering force has been observed in Ref.111 and would require an additional force dependence of the ATP dissociation constant (that increases with external force).

4.3.1.4 ATP hydrolysis versus kinesin stepping

To determine the coupling ratio between ATP hydrolysis and kinesin stepping, the average number of ATPs hydrolyzed for every kinesin step (backward or forward) was calculated as a function of external force. It is seen from Fig. 4.5 that the chemomechanical coupling ratio is close to unity in a large range of external forces.

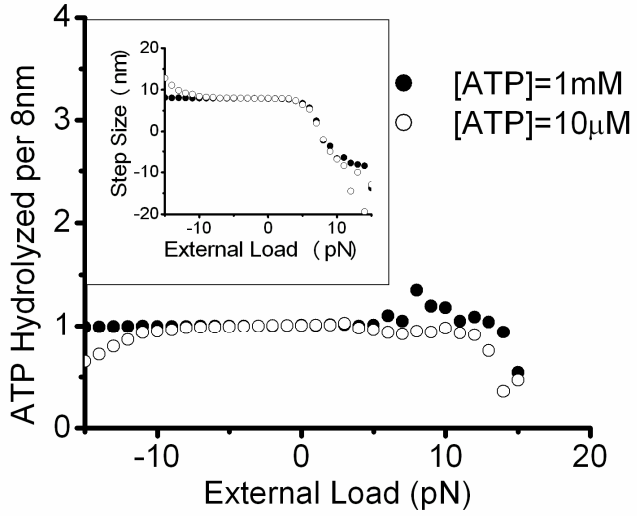


Figure 4.5. The chemomechanical coupling ratio and average step size as a function of external force.

The higher the ATP concentration, the larger is the force range that the tight coupling sustains (−15 to 6 pN for [ATP] = 1 mM and −10 to 6 pN for [ATP] = 10 μM). At large negative forces and a low concentration of ATP, kinesin takes more than one 8-nm step per ATP hydrolyzed, corresponding to a forced sliding in the plus direction. The forced sliding is not seen for [ATP] = 1 mM and forces less negative than −15 pN,

suggesting a stronger binding of kinesin to the microtubule in the presence of higher concentrations of ATP. The force dependence of chemomechanical coupling ratios at large positive forces shows a more complicated behavior in the region of the forward/backward stepping transition: Near the stall force, the kinesin stepping becomes loosely coupled to ATP hydrolysis, with more than one ATP consumed for a successful step. Taking into account that the net motion is close to zero in this range of force, this loose coupling is due to the balance between the force produced by kinesin and the external load, which leads to a quick forward/backward motion of one head without moving the other. However, when the force further increases, the chemomechanical coupling ratio approaches unity again, indicating the recovery of a tight coupling during the consecutive backward motion in this range of external load. This ratio becomes <1 at even larger hindering forces (>15 pN for [ATP] = 1 mM), and kinesin slippage to the minus end of the microtubule occurs (Fig. 4.1 d).

4.3.2 On the asymmetry of kinesin walking

4.3.2.1 On the limping of kinesin

In the experiment of Block and coworkers, they observed that the truncation of the central stalk induces kinesin limping: during the kinesin walking, the dwell times consist of alternating short and long intervals.[118] The limping factor L was defined as the ratio between the successive long and short dwell times in order to evaluate the limping degree.[118] As proposed in the present model, the asymmetry of the kinesin walking is induced by the length change of neck-linkers. The experimentally observed relation between the truncation length and the limping factor (and dwell time) is very similar to

the calculated neck-linker length dependence of the limping factor in the present model. Fig. 4.6 demonstrates the influence of the neck-linker length to the calculated limping factor as well as the dwell time: the mean long dwell time increases largely with the shortening of the "leg length" *d* whereas the mean short dwell time remains invariant (Fig. 4.6 (a)); As a result, the limping factor increases with the shortening of the leg length *d* (Fig. 4.6 (b)). The calculated limping factor with neck-linker lengths of 4.6 nm in unwound and 4.1 nm in rewound state is similar to that obtained in the experiment for truncated kinesin DmK401[118], which has the shortest stalk in the series of tested kinesin constructs (see Fig. 3(B) in Ref.118).

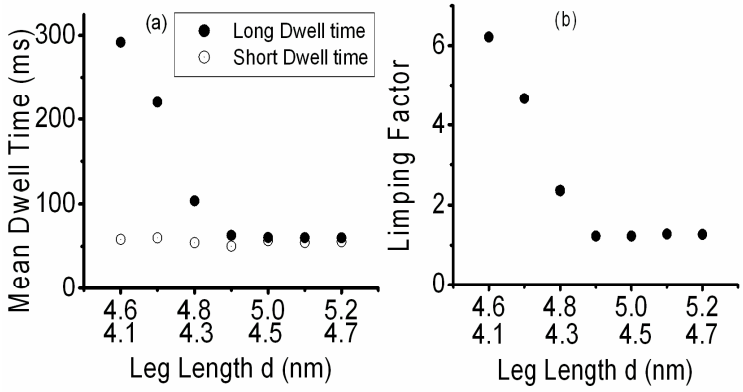


Figure 4.6. (a) The neck-linker length dependence of dwell times of the two kinesin heads. The solid and hollow points are for the longer and shorter dwell times, respectively. (b) The neck-linker length dependence of the limping factor of the kinesin. F_{ext}= 4 pN, and [ATP]=2 mM.

We calculated the trace of the center of the kinesin with the shorter neck-linker lengths of 4.6 nm and 4.1 nm (truncated kinesin) which is shown in Fig. 4.7 (a). One observes the alternating short and long intervals between forward steps.

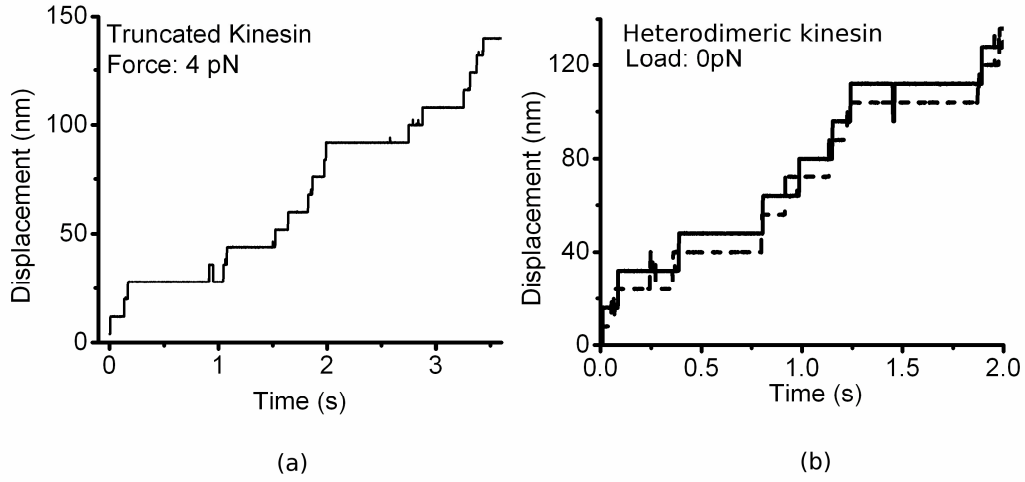


Figure 4.7. (a) The trajectory of the center of a kinesin with neck-linkers of 4.6/4.1 nm in the presence of a hindering force of 4 pN. (b) The trajectory for the motion of the two heads of the heterodimeric kinesin in the absence of external load. [ATP]=1 mM

Figure 4.8 (a) demonstrates the distribution of the long and short dwell times and Fig. 4.8 (b) shows the distribution of the individual limping factor. As seen from the figures, there is a large difference between dwell time distributions of the fast and slow phases. The mean dwell times for the slow and fast steps are $t_{slow}\approx290$ ms and $t_{fast}\approx58$ ms, respectively. The mean value of the limping factor L is about 6.22. These values are

consistent with the experimental results from Block and coworkers (t_{slow}≈200 ms, t_{fast}≈40 ms and L=6.45±0.31, see Fig. 3(B)&(C) in Ref. 118).

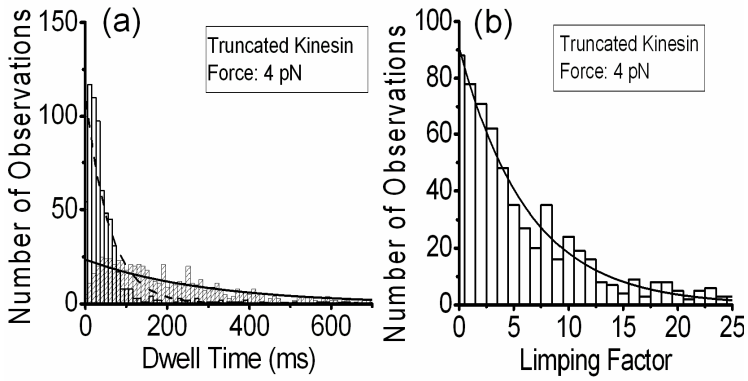


Figure 4.8. (a) Dwell times of the two kinesin heads. The hollow and filled bars are for the fast and slow steps, respectively. (b) The distribution of the limping factor for a kinesin with neck-linkers of 4.6/4.1 nm. F_{ext}= 4 pN, and [ATP]=2 mM.

As a comparison, one can see from Fig. 4.9 (a) that dwell times of fast and slow phases of wild-type kinesin (which as mentioned earlier have longer neck-linkers) are very close: t_{slow}≈56 ms and t_{fast}≈60 ms. The corresponding mean value of limping factor (Fig. 4.9 (b)) is about 1.27, confirming that there is no limping during the walking of the wild-type kinesin.

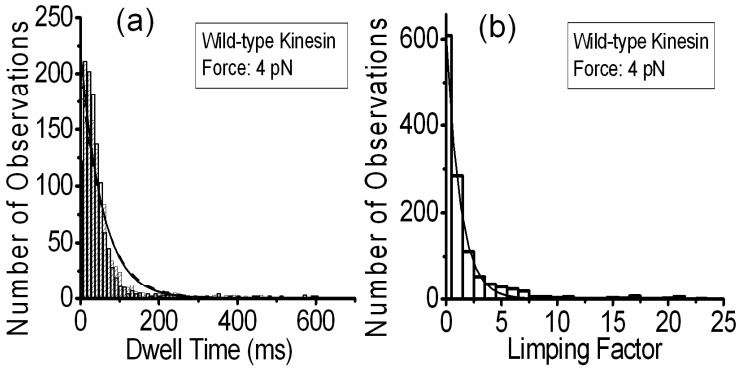


Figure 4.9. (a) Dwell times of the two kinesin heads. (b) The distribution of the limping factor for a wild-type kinesin (with neck-linkers of 5.2/4.7 nm). F_{ext}=4 pN, and [ATP]=2 mM.

We also calculated the walking speed of this truncated kinesin as a function of external force with [ATP] = 1mM and [ATP] = 10 μM. As shown in Fig. 4.2 (b), the force dependence of the walking speed of this kinesin is very similar to that of the wild-type kinesin, although the speed of the former is slower. We calculated the limping factor under various external forces. As shown in Fig. 4.10, kinesin limping becomes less severe at both large assisting and hindering forces, although the force at which the limping factor takes a maximum value depends on the system parameters (in particular the force dependence of the chemical reaction rate constants), and further experiments are needed to determine these parameters.

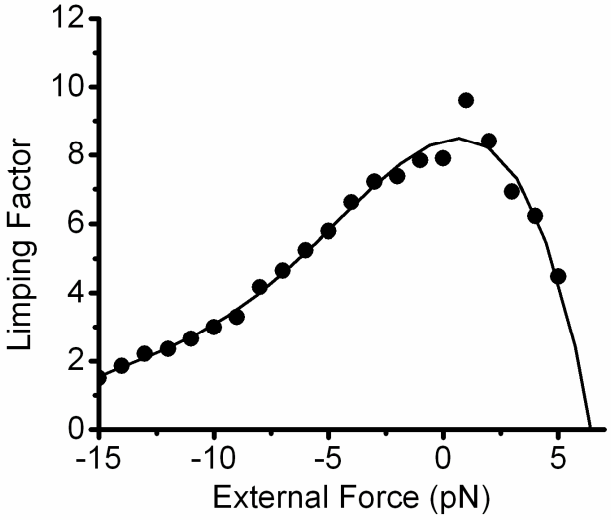


Figure 4.10. The predicted force dependence of the limping factor of the kinesin with neck-linkers of 4.6/4.1 nm.

In addition, we calculated the average number of ATP molecules hydrolyzed per kinesin walking step to determine the chemomechanical coupling ratio for the truncated kinesin. Figure 4.11 (a) demonstrates the calculated coupling ratio under different forces. Over a large range of forces (–15~12 pN for [ATP] = 1 mM and –5~5 pN for [ATP] = 10 μM), the coupling ratio is close to unity, which means only one ATP molecule is consumed for each 8-nm step. Out of the force range mentioned above, the ratio decreases with the increase of the magnitude of the external force, corresponding to the sliding of kinesin induced by large external forces.

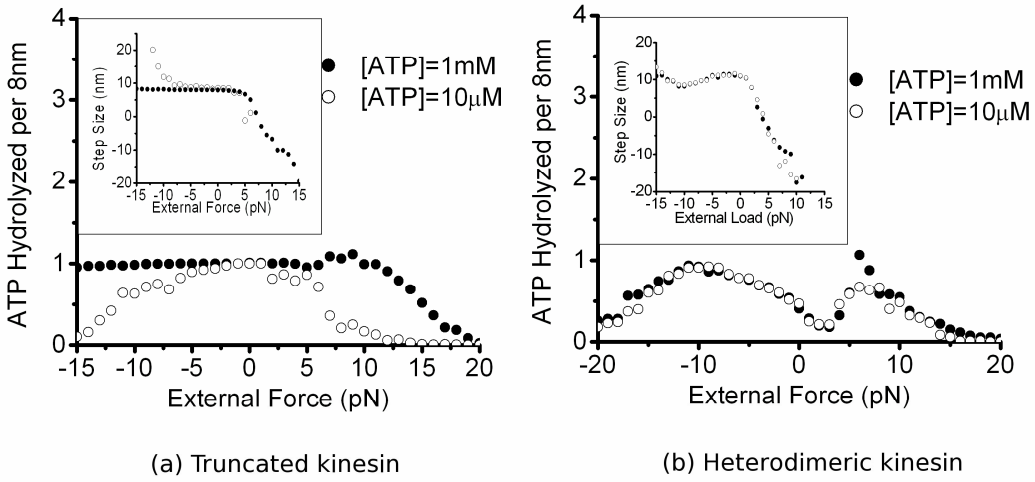


Figure 4.11. The predicted force dependence of the number of ATP molecules hydrolyzed per 8-nm movement of (a) the kinesin with neck-linkers of 4.6/4.1 nm and (b) the heterodimeric kinesin mutant.

4.3.2.2 Heterodimeric kinesin mutant

In Fig. 4.7 (b), a typical trajectory is shown for the walking of a heterodimeric kinesin mutant, for which the hydrolysis activity of one of the two heads is removed, in the absence of external force. It is a hand-over-hand motion: during each step one head moves 16 nm while the other head keeps attached to the microtubule, the same as during the stepping of a wild-type kinesin. The speed of the mutant kinesin obtained at a saturating ATP concentration of 1 mM is ~80 nm/s, which is ~6-fold smaller than that of wild-type homodimer (~450 nm/s). This result is consistent with the experimental data by Thoresen and Gelles: A biotinated Drosophila kinesin construct (K401-BIO-H6) with a switch I domain point mutation walks with a speed of ~90.3 nm/s at 1 mM of ATP.[120]

'Limping' is easily observed from the trajectory, although this limping is by a mechanism that is different from the one of the truncated kinesin.

We calculated the mean values of the long and short dwell times as well as the limping factor of the heterodimeric kinesin mutant under various external forces. The results are shown in Fig. 4.12.

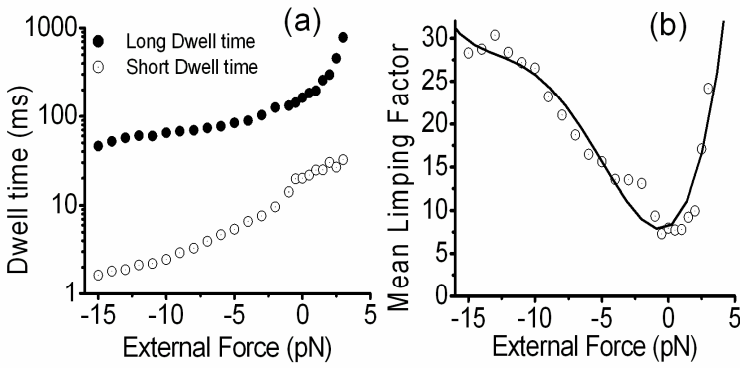


Figure 4.12. (a) The predicted force dependence of the average long/short dwell times of the heterodimeric kinesin mutant with [ATP] = 1 mM. (b) The predicted force dependence of the limping factor.

One can observe from Fig. 4.12 (a) that the increase of the hindering force prolongs the mean dwell time of both fast and slow phases, whereas an assisting force makes the dwell times in both phases shorter. The hindering force has a stronger influence on the long dwell time, while the assisting force affects more significantly the short dwell time.

As a result, the difference between the long and short dwell times and the limping factor (the ratio of the mean value of long and short dwell time) reaches its minimum at around 0 force (as shown in Fig. 4.12 (b)).

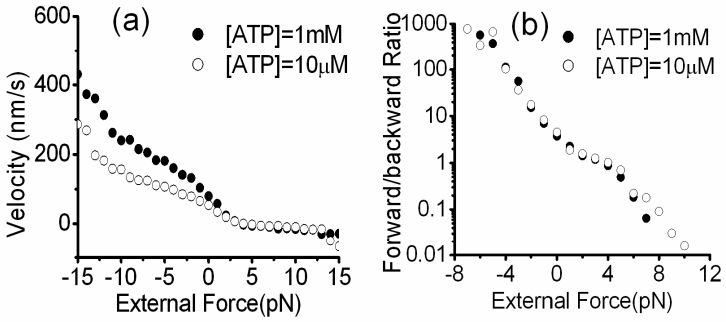


Figure 4.13. (a) The predicted force dependence of the walking speed of the heterodimeric kinesin mutant, with [ATP] = 1 mM (filled symbols) and [ATP] = 10 μM (open symbols). (b) The predicted force dependence of the forward/backward step ratio of the heterodimeric kinesin.

The predicted force dependence of the walking speed of the kinesin mutant is shown in Fig. 4.13 (a). The mutant and wild-type kinesins differ in that the speed of the former increases monotonically with the assisting force, while the latter has a maximum speed at ~–5 pN. Figure 4.13 (b) shows the force dependence of the forward/backward step ratio at ATP concentrations of 1 mM and 10 μM. As seen from this figure, the influence

of the ATP concentration on the forward/backward ratio is very small, and in both cases the ratio decreases monotonically with the external force. The ratio becomes unity when the force is about 3.5 pN, corresponding to a stall force that is about half of that of the conventional kinesin.

We also calculated the mean value of the number of ATP molecules consumed for each kinesin mutant step. As seen from Fig.4.11 (b), the number of ATP molecules hydrolyzed per 8-nm step is around 0.5 under near vanishing forces, in accordance with the fact that ATP hydrolysis occurs in only one head and the observation that the hydrolysis of one ATP couples to a forward step of each of the two heads.

4.3.2.3 Stepping of wild-type kinesin in the presence of both ATP and AMPPNP

We calculated the motion of the wild-type (conventional) kinesin in the presence of 2 mM ATP and 0.2 mM AMPPNP. A segment of the trajectory for the motion of the center of kinesin is shown in Fig. 4.14. The trajectories given in that figure were obtained in the absence and in the presence of a 3 pN hindering force, respectively. It is seen clearly that in the presence of the hindering force (Fig. 4.14 (a)), the kinesin frequently takes long pauses. These long pauses are normally terminated by a quick backward step, which is followed by either another segment of long pause or by forward steps. The average waiting time before those backward steps is ~3 seconds. On the contrary, the long pause during the processive motion of the kinesin appears rarely at small external forces (as shown in Fig. 4.14 (b)). These calculated results are again consistent with experiments: It was observed that in the presence of both ATP and AMPPNP, the conventional kinesin still walks processively. However, under a hindering

force (~5 pN), the processive forward movement is halted frequently by long waiting periods and the forward motion is reactivated by an obligatory backward step.

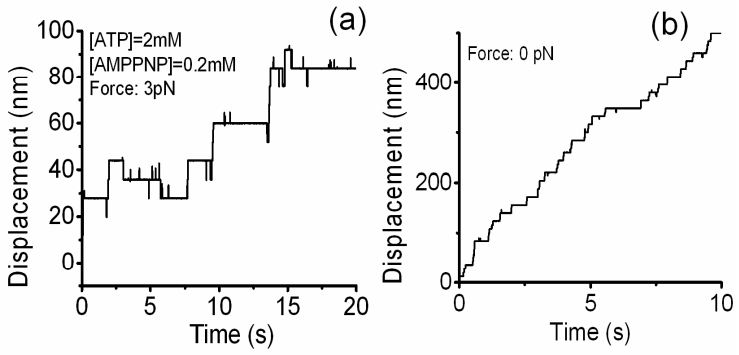


Figure 4.14. A typical trajectory of the motion of the conventional kinesin in the presence of 2 mM ATP and 0.2 mM AMPPNP in the presence of a load of 3 pN (a) and in the absence of any load (b).

The long waiting period, in our model, is induced by a state of the kinesin in which the front head is occupied by ATP or ADP/Pi and the rear head is occupied by AMPPNP. As we assume in the model,[167] the ATP binding as well as the following ATP hydrolysis in the front head supplies a driving force. The rear head which is bound with AMPPNP is detached from the microtubule and pulled forward if no external hindering force is applied on the kinesin, due to the stronger binding and larger driving force generated by ATP binding and hydrolysis in the other head. On the other hand, a large hindering force

will balance the forward driving force on the rear head and therefore prohibit the rear head from moving forward. This stalled state survives a long time until the front head is drawn backward under the hindering force. Therefore, the two heads switch the leading position. The AMPPNP on the new front head is now released. If ATP binds to the head, kinesin resumes walking forward: the hydrolysis product Pi in the new rear head is released, the ADP-bound head is easily detached from the microtubule under the driving force generated by the ATP-bound front head. However, if AMPPNP binds instead, the kinesin will be able to take one single forward step and then go back to the stalled state with an AMPPNP bound to the rear head.

4.4 Conclusion

In the present study, we present a simple model[166,167] to simulate the asymmetric hand-over-hand translocation of the conventional kinesin and its mutants (the truncated kinesin with shorter neck-linkers and the heterodimeric kinesin mutant) on the microtubule. The asymmetry of the kinesin walking is induced in the present model by the neck-linker change due to the rewinding and unwinding of the lower end of the coiled-coil during the kinesin walking.[167] The main goal of the present work is to understand the recent single molecule experiments on the mechanical response of the kinesin and its mutants to external loads[116,118,120] and to construct a simple theoretical model that will allow us to understand forward and backward motions of kinesins and to make experimentally testable predictions.

We have studied the force dependence of both forward and backward stepping of the conventional kinesin. The model perfectly mimics a hand-over-hand mechanism for

kinesin in both small and large forces (−15 to +15 pN for 1 mM ATP), and for motions in both minus and plus directions, given that the ATPase activity of the kinesin head depends only on its conformation. However, when a very large force is applied, one could observe an inchworm walking/sliding mechanism (Fig. 4.1 (*d*)). It was found that the force up to which the hand-over-hand mechanism sustains decreases when the ATP concentration decreases.

In this hand-over-hand mechanism, the prestroke state consists of one ADP-occupied and one empty kinesin head. In this prestroke state, the empty head takes the leading position (to the plus direction). Both kinesin heads are bound to the microtubule, and the front one binds more strongly. This feature of the current model is different from that proposed by Carter and Cross[116] and, as shown in Scheme 4.1, is consistent with the observed 17- and 0-nm alternate steps for each head.[112] ATP binding to the front empty head induces a conformational change of its neck-linker, which prefers a tilting angle of ≈45° and thus prefers to be the rear head through the interaction between the neck-linkers connected to the heads[94,98] (see Scheme 4.1). In the absence of a large hindering force, the backward movement of this head is very unlikely because of its strong binding to the microtubule. In contrast, the rear ADP head binds the microtubule weakly and is easily detached. It is thus pulled forward by the front head (bound with ATP) until it reattaches to the next microtubule-binding site, which is ~16 nm away from its original binding site. This motion makes the original front head (with an ATP bound) a rear one, the neck-linker of which therefore redocks to its stable conformation (see Scheme 4.1 (a), states B–D). The reattachment of the ADP-occupied head further decreases its ADP-

binding affinity and allows ADP to be released, whereas ATP hydrolysis and P_i release (both believed to be fast) transform the rear head to the ADP state, which is only weakly bound to the microtubule. Therefore, in the forward motions, through each of the reacting cycles, one of the two heads moves to the plus end of microtubule by ~16 nm, and the other head stays tightly bound to its microtubule-binding site. ATP binding to the new empty head initiates another cycle of ATP hydrolysis and therefore the next forward step of kinesin, with the two heads exchanging their roles from the previous step. As a result, one observes the hand-over-hand walking mechanism of kinesin.

The model was further used to investigate the influence of the external force on the walking mechanism of the conventional kinesin. Consistent with experiments, the hindering forces not only slow down the forward motion of kinesin but also lead to more frequent backward steps, thus decreasing the forward/backward ratio. The backward stepping occurs by the following mechanism: ATP binding to the front head generates a force between the two heads through the neck-linkers so that the front head is pulled in the backward direction and the rear head is pulled to the plus direction. When the external force is small, it is the weakly bound rear head that detaches, as discussed earlier. A large hindering force, however, balances with the forward force on the rear head that is generated from the front head and prevents the forward motion of the rear head (Scheme 4.1 (a), state C). Therefore, a large enough hindering force holds the rear head at its microtubule-binding site. The front head with an ATP bound, however, is subject to the power stroke and the external force, both in the minus-end direction (Scheme 4.1 (a), state C). These two forces together detach the front head in the ATP

state from the microtubule and drive its motion in the minus direction. When this detached head reattaches a microtubule-binding site that is ~16 nm away from its original binding site, in the minus direction, the driving force due to the power stroke vanishes, and it binds strongly to the microtubule until another ATP binds. As a result, consecutive backward steps are achieved.

What is somewhat surprising in the force dependence of the motion of the conventional kinesin is that the assisting force does not always increase the speed of kinesin. In fact, as mentioned earlier, in the force range of ≈−5 to −15 pN, the larger the assisting force, the slower the motor walks (see Fig. 4.2). The other surprising result, again consistent with experiments, is that the dwell time of forward steps and the dwell time of backward steps both increase with the applied force and that the forward and backward stepping have the same dwell times. These two findings are both consistent with experiments and are easily understood given the stepping mechanism discussed above.

Although the assisting force increases the speed of the physical translocation of kinesin, it also displaces the kinesin heads slightly from their optimal binding positions on the microtubule and slows down the chemical transitions. The observed decrease of kinesin speed with the increase of relatively large assisting force suggests that the rate-limiting steps are chemical rather than physical. Consistently, experiments have shown that the time that is required for a conventional kinesin to take a physical step is in the range of 10 μs for both forward and backward steps,[116] much shorter than the overall turnover time (on the order of 10 ms or slower). This observation of the chemical

transitions being the rate-limiting steps is also consistent with the force dependence of the dwell times. As shown in Fig. 4.5, the hindering force has a large effect on the dwell time: The dwell time increases by a factor of ≈20 when the external force increases from 5 to 10 pN ([ATP] = 1 mM). This effect is the same for the forward and backward steps.

The increase of the dwell time of the forward step in the presence of an opposing force is in accord with the lower frequency of its occurrence. However, increase of the dwell time of the backward steps accompanies the increase of its frequency of occurrence. Because in both forward and backward stepping described above ATP hydrolysis occurs in the rear head and ADP releases from the front head (to the plus end), the dwell times of both forward and backward motions show the same force dependence. Therefore, the applied force in the minus direction decreases the overall turnover rate of ATP hydrolysis, and the chemical transitions remain as the rate-limiting steps. Because the chemical states at which the forward and backward stepping take place are the same, the dwell times of forward and backward steps are the same and respond to the external load in the same way.

The force dependence of the conventional kinesin discussed above makes it a tightly coupled motor in a large range of external loads. This tight coupling is achieved because in both forward and backward walking, ATP binding to an empty site is required for the detachment of one of the two heads. And in a large range of forces, one of the two heads (in the presence of small forces and forward stepping, it is the front head, and in backward stepping, it is the rear head) binds strongly to the microtubule. Therefore, both forward and backward steps are tightly coupled to ATP turnover. It is because of this

tight coupling that the stall force does not depend on the ATP concentration sensitively. These properties of kinesin make it a strong motor with high efficiency when it walks against large forces and ensure that kinesin can sustain a large negative force. Even when the force is much larger than the stall force (>10 pN), kinesin steps backward with a slow speed in an ATP-dependent manner, instead of sliding backward quickly or detaching from the microtubule (see Fig. 4.2).

The translocation of the conventional kinesin on the microtubule doesn't display the asymmetry character. To understand the asymmetry of the hand-over-hand mechanism, we studied the processive walking of two kinesin mutants by using the same theoretical model. The first is a truncated kinesin homodimer with shorter neck-linkers (used as a possible model for the kinesin with a truncated stalk) and the second is a kinesin heterodimer, one head of which is deprived of ATP hydrolysis activity, although it does bind ATP. Results from our model shows that both mutant kinesins limp during walking. For the truncated kinesin, Asbury *et al.* speculated that the limping is due to the change in the neck-linker length during the stepping of kinesin heads[118]. As described in the asymmetric hand-over-hand mechanism, the lower portion of the coiled coil unwinds and rewinds in the successive steps. As a result, conformations of kinesin heads (represented in this model by the angles between the neck-linker and the microtubule, see Scheme 4.1 (b)) are different depending on which head takes the front position. Chemical transitions, including ADP release, ATP binding, and ATP hydrolysis, are sensitive to the conformations of kinesin heads.

The neck-linkers of a conventional kinesin are long enough that the length change due to the unwinding or rewinding of the coiled coil has little influence on the conformations of its bound heads. On the other hand, the shortening of the central stalk is assumed to result in the shortening of neck-linkers, e.g., through misregistration of the coiled coil, by ~1 nm.[118] The shortened neck-linker imposes more severe constraints on the conformations of kinesin heads when they are bound to the microtubule (e.g., angles between the neck-linker and the microtubule deviate from their optimal values). The larger the deviation of the head conformation from its chemical transition conformation, the slower the chemical transition (see Scheme 4.1 (b)). As a result, the unwinding and rewinding of the coiled coil have a more apparent effect on the chemical transitions with this shortened neck-linker: Only when the central stalk unwinds does the kinesin conformation allow fast chemical transitions.

The dwell time of the conventional kinesin is ~60 ms under a hindering force of 4 pN, while for the kinesin with shorter neck-linkers under the same force, the dwell time is split into alterative fast and slow phases. The fast dell time (with an unwound coiled coil) is changed only slightly compared to the dwell time of the conventional kinesin, while the slow interval (with a misregistered coiled coil) becomes ~5 times longer. As mentioned earlier, the length of the neck-linker (plus that of the kinesin head) is represented by the parameter d. Values of 5.2 and 4.7 nm were used for unwound and rewound the conventional kinesin, respectively. No limping is observed in the model with these two values. However, with the synchronous decrease in the two values, the limping becomes more and more evident (Figure 4.7 (a)). The limping factor reaches

6.22 with the values of 4.6 and 4.1 nm for a shortened kinesin, which is very close to the experimental value of kinesin DmK401.[118]

We use a construct similar to the conventional kinesin to calculate the heterodimeric kinesin mutant with one head lacking the ability to hydrolyze ATP, which was also shown in our model to walk by the asymmetric hand-over-hand mechanism. Assuming that Head 1 does not hydrolyze ATP, we start with the pre-stroke state, in which Head 2 is occupied with ADP and Head 1 is empty. Head 1 takes the leading position. Binding of ATP to the front empty head (leads to an empty to ATP state transition) induces a conformational change in the neck-linker and exerts a driving force that detaches the rear head and moves it forward. The difference between the mutant and the conventional kinesin is that Head 1 does not hydrolyze ATP and thus generates a smaller driving force [it is assumed in the model that in addition to ATP binding, the hydrolysis of ATP to ADP/P_i further provides driving force for kinesin walking, in accordance with fluorescence studies which show that ATP hydrolysis induces further kinesin conformational changes[100,168]].

Without a large hindering force, the rear head occupied by an ADP is easily detached from the microtubule and pulled forward by the front head. After Head 2 rebinds to its new microtubule binding site that is 16 nm from its previous binding site, ADP releases from it. The two heads have switched their leading positions. Although Head 1 does not hydrolyze ATP and binds the microtubule relatively strongly, ATP binding and hydrolysis in Head 2 generate a stronger binding state and a sufficiently large driving force to detach the ATP-occupied Head 1, and the competition between

these two heads leads to a forward motion of the more weakly bound ATP-occupied rear head while Head 2 remains bound. This step is slower than the previous step as a result of the binding of an ATP (or AMPPNP)-occupied kinesin head that is much stronger than that of an ADP-occupied kinesin head to the microtubule. Consequently, the mutant kinesin limps when it walks. The resulting short dwell time corresponds to the first step: the front head is occupied by ATP, which remains in the pre-hydrolytic state, and the rear head is ADP-bound. The long dwell time, on the other hand, corresponds to the state in which the front head is occupied with the post-hydrolytic ADP/P_i and the rear one is occupied by ATP. Because of the weak binding of the ADP-bound head to the microtubule, the assisting force will shorten the short dwell time more significantly than the long dwell time. The limping factor increases with an increase in the assisting force. On the other hand, the small hindering force has little influence on the short dwell time, whereas it aggravates the difficulty for the forward step to take place in the latter state. The limping factor also increases with the increase in hindering force.

It is seen from the description given above that the hydrolysis of one ATP molecule generates two successive 8 nm steps for a mutant kinesin. As a result, the stall force is smaller for a mutant kinesin (~3.5 pN) than for the conventional kinesin (~7–8 pN) and the former walks slower than the latter. The smaller stall force and the slowness of the mutant kinesin could also be understood from the force-generating mechanism. In the conventional kinesin, in every step a driving force generated from ATP binding and hydrolysis is used to detach the ADP-occupied kinesin head, while in a mutant kinesin,

every other step the force is used to detach an ATP-occupied kinesin head, which is much more strongly bound to the microtubule.

The simple theoretical model we used for the conventional kinesin, truncated kinesin with shorter neck-linkers, and heterodimeric kinesin mutant[118,120] provides possible explanations for the various experiments. As discussed above, although both the truncated homodimer and the mutant heterodimer exhibit limping behavior, the origins of limping are different. The chemomechanical coupling ratio of the truncated homodimer is the same as that of the conventional, and in the presence of small external forces, the hydrolysis of every ATP molecule leads to one 8 nm step whereas the consumption of one ATP molecule induces two consecutive 8 nm steps for the mutant heterodimer. The force dependence of motion, including the stall force of the truncated kinesin, is also similar to that of the conventional kinesin. Although both reconstructed kinesins walk at a similar speed in the absence of the external force, the heterodimer shows a much stronger force dependence and its stall force is only approximately one-half of that of the conventional kinesin and the truncated kinesin.

CHAPTER V

CONFORMATIONAL CHANGE OF KINESIN STUDIED BY ANISOTROPIC NETWORK MODEL

5.1 Introduction

In Chapter IV we studied the processive motion of the dimeric kinesin on the microtubule by using a simplified model in which the two kinesin heads are treated as two points and their movement traces are described by the Langevin equation.[166,167] It is proposed in that model that ATP binding (as well as the following ATP hydrolysis) to the front kinesin head results in a large-scale conformational change of the neck-linker and therefore generates a driving force for the kinesin movement. This conformational change of the neck-linker has been detected in earlier experiments.[91-95] It was also observed in the single molecule experiment that there is local conformational change in the nucleotide-binding domain when ATP binds to the kinesin head.[90,169] Therefore it is reasonable to believe that for the dimeric kinesin there is correlation between the structural change of the nucleotide binding site and the remote neck-linker: the structure of the nucleotide binding site is changed after the ATP binding and this structural change transmits to the remote neck-linker and eventually leads to a large global conformational change of the neck-linker. In contrast, the monomeric kinesin deprives the ability of walking on the microtubule processively although ATP can still bind to the head and produce the similar structural change around the nucleotide binding site. One can speculate that the structural correlation between the nucleotide binding site and the neck-linker might be deprived in the monomeric kinesin.

Here we simulated the global motions of the kinesin using the coarse-grained normal mode analysis, the anisotropic network model (ANM) and then studied the solvent effect on the kinesin motions using Langevin mode analysis. To investigate the correlation between the nucleotide binding site and the neck-linker, we calculated the time-average cross-correlation function between the corresponding residues for both dimeric and monomeric kinesin.

5.2 Theoretical Model

5.2.1 Anisotropic network model (ANM)

Elastic network model has a fundamental assumption that a folded protein can be regarded as a large elastic network linked by the nodes (C_α atoms).[50] All nodes are connected by uniform springs which share the same force constant. Interactions between C_α's are limited by a defined cutoff distance r_c. Within the cutoff distance the potential function of a protein system with N residues is expressed as:[50]

$$V = \frac{\gamma}{2}\sum_{i}^{3N}\sum_{j}^{3N}(r_{ij} - r_{ij}^{0})^2; r_{ij} < r_c \tag{5.1}$$

where γ is the force constant of springs and r_{ij} is the distance vector between i^{th} and j^{th} C_α atoms. Atilgan and co-workers examined the influence of the parameter values of γ and r_c on the generated normal modes in ANM by comparing the theoretically calculated mean-square fluctuations to the experimental B-factors in many protein systems[50]. The best-fittings of B-factors were obtained in the range that the force constant is 1.0±0.5 kcal/(mol·Å^2) and the cutoff distance is within12-15 Å.[50]

The formula of the potential energy function (Eq. 5.1) determines that ANM model could predict not only the amplitude but also the direction of the fluctuation motions.

The motion direction in each normal mode is provided by 3N-dimensional eigenvectors obtained from the diagonalization of the second derivative matrix of the potential function (Hessian matrix, *H,* corresponding to F_{ij} in Eq. 1.9),

$$H = U\Lambda U^{-1} \tag{5.2}$$

where Λ is a diagonal matrix containing eigenvalues of H matrix, the elements of which are equal to the square of normal mode frequencies ($\Lambda_{i\alpha}=\omega_{i\alpha}^2$, $(i\alpha)=1,\ldots,3N$, where $i=1,\ldots,N$ is the atom index and $\alpha=1,2,3$ the motion directionality), U is an orthogonal matrix containing eigenvectors of H organized by column. The first six extremely small normal mode frequencies correspond to the translation and rotation and therefore are not involved in the data analysis. The unit of the normal mode frequency ω received from matrix operation above is radian. It is converted to unit expressed by wave numbers (cm^{-1}) with coefficient 108.59 ($1cm^{-1}$ =108.59radian).[128]

The equilibrium correlation function reveals the important information such as the fluctuation of protein structure and the correlation of motions of different parts of the protein. The function is given by the following formula,

$$\left\langle \Delta r_{i\alpha} \bullet \Delta r_{j\alpha} \right\rangle = \frac{k_B T}{\gamma} \sum_{k=1}^{3N-6} \frac{U_{(i\alpha)k} U_{(j\alpha)k}}{\omega_k^{\,2}} \tag{5.3}$$

Thermal B-factor, which could be experimentally measured and used as a main criterion for normal mode analysis, can be obtained from self-correlation function by the following relationship, $B_i = \frac{8\pi}{3} \sum_{\alpha=1,2,3} < \Delta r_{i\alpha}^{\,2} >$.

5.2.2 Normal modes from ANM

A great advantage of normal mode analysis is that a single (or sometimes a few of) low-frequency mode can describe the functional large-scale conformational change of protein despite the fact that the protein is simulated in vacuum in ANM whereas the real protein dynamics or structure changes occur in solution. ANM, as a coarse-grained model with simple potential function, has been applied to hundreds of protein systems to testify the overlap of normal modes and the relevant conformational change transitions and even used to predict conformational changes in the case that only one initial state crystal structure is valid.[137,138,140-142]

In ANM, the calculated eigenvector $U_{(i\alpha)(j\beta)}$ provides the motion directionality of j^{th} C_α atom in the direction of *x, y* or *z* (β=1,2,3 respectively) under the specific $(i\alpha)^{th}$ normal mode. Given the initial coordinates and the specific normal mode, the trajectory of j^{th} C_α atom can be calculated by disassembling the corresponding eigenvector, as shown in the following equation:

$$r_{j\beta} = r_{j\beta,0} + \kappa_{i\alpha} \frac{U_{(i\alpha)(j\beta)}}{\sqrt{\sum_{k=1}^{3N-6} U_{k(j\beta)}^{\ 2}}} \cos(\omega_{i\alpha} t) \quad ,\alpha,\beta = 1,2,3 \tag{5.4}$$

where $\kappa_{i\alpha} = (\frac{2k_B T}{\omega_{i\alpha}^{\ 2}})^{1/2}$ is the motion amplitude. The trajectory of each residue is generated and then analyzed using the software VMD.

5.2.3 Langevin mode analysis

In ANM, 3N-6 non-zero normal modes are obtained. Each normal mode represents a characteristic motion of the whole protein. A few of the lowest-frequency normal modes

give good description of the biologically functional dynamics of the protein under study.[139,140] Therefore the solvent effect on the protein dynamics could be taken into account by focusing on the low-frequency normal modes. In this section we describe the details of Langevin mode analysis proposed by Lamm and Szabo.[144] Here, we use coarse-grained model to treat a protein as a large elastic network and each residue is chosen to be a node, the same as in ANM. The dynamics of the N residue units (nodes) in total 3N degrees of freedom can be described by Langevin equation,

$$\ddot{q}_{i\alpha} + \sum_{j\beta=1}^{3N} \zeta_{(i\alpha)(j\beta)} \dot{q}_{j\beta} + \sum_{j\beta=1}^{3N} \frac{V''}{\sqrt{m_i m_j}} \left(q_{j\beta} - q_{j\beta}^0 \right) = R_{i\alpha}(t), \qquad (i\alpha) = 1,...,3N \tag{5.5}$$

where i, j (=1,…,N) denote the atom indices, α, β (=1,…,3) is the degree of freedom, $q_{i\alpha}$ is the Cartesian coordinate describing the position of i^{th} node in 3-dimensional space and $q^o_{i\alpha}$ is the equilibrium position, m_i denotes the mass of i^{th} node, ξ denotes the mass-weighted friction matrix and V'' denotes the second derivative matrix of potential energy. $R_{i\alpha}(t)$ is the stochastic force from the surrounding environment. Note that in the absence of any solvent (ξ=0, $R(t)$=0), Eq. 5.5 is simplified to Newton's equation. The solution of the simplified equation is the 3N normal modes from the eigenanalysis of mass-weighted Hessian matrix H ($H_{(i\alpha)(j\beta)} = V''_{(i\alpha)(j\beta)} / \sqrt{m_i m_j}$).

The friction matrix is a symmetric 3N×3N matrix with the diagonal submatrices (3×3) representing the friction constant on each node for each degree of freedom and the off-diagonal submatrices representing hydrodynamic interactions between nodes. Without taking into account the inter-node hydrodynamic interactions, Stokes's law provides the friction constant in the solution by a diagonal matrix with the elements:

$$\gamma_{i\alpha}=6\pi\eta\alpha_i/m_i \tag{5.6}$$

where η is the viscosity of solvent, α_i is the hydrodynamic radius of i^{th} node. The hydrodynamic radius α_i can be estimated from the solvent accessible surface, s, van del Waals radius of each node, r_{vdw} and the radius of solvent molecule, r_p, as shown in the following equation,[144,170]

$$\alpha_i = r_{vdw}\sqrt{\frac{s_i}{4\pi(r_{vdw}+r_p)^2}} \tag{5.7}$$

However, hydrodynamic interactions of inner-nodes should be considered when nodes shield with each other from the solvent effects. Therefore the definition of Oseen tensor T was introduced and the friction matrix is obtained by:[144,171]

$$\gamma_{(i\alpha)(j\beta)} = \frac{(T^{-1})_{(i\alpha)(j\beta)}}{\sqrt{m_i m_j}} \tag{5.8}$$

The lowest order formation of Oseen tensor is expressed by:

$$\begin{aligned} T^0_{(i\alpha)(j\beta)} &= (6\pi\eta\alpha_i)^{-1}\sigma_{(i\alpha)(j\beta)}, i=j \\ T^0_{(i\alpha)(j\beta)} &= (8\pi\eta|r_{i,j}|)^{-1}\left[\sigma_{\alpha\beta}+\frac{(r_{i,j}r_{i,j})_{\alpha\beta}}{|r_{i,j}|^2}\right], i\neq j \end{aligned} \tag{5.9}$$

where $r_{i,j}$ represents the distance vector between i^{th} and j^{th} nodes.

The extended formation of Oseen tensor was proposed by Rotne and Prager for the case of large hydrodynamic radii:[144]

$$\begin{aligned} T^{RP}_{(i\alpha)(j\beta)} &= T^0_{(i\alpha)(j\beta)}, i=j \\ T^{RP}_{(i\alpha)(j\beta)} &= T^0_{(i\alpha)(j\beta)} + (8\pi\eta|r_{i,j}|^3)^{-1}(\alpha_i^2+\alpha_j^2)\left[\frac{1}{3}\sigma_{\alpha\beta}-\frac{(r_{i,j}r_{i,j})_{\alpha\beta}}{|r_{i,j}|^2}\right], i\neq j \end{aligned} \tag{5.10}$$

These equations above are available only if no nodes overlap with each other ($r_{i,j} > \alpha_i+\alpha_j$). In the case that there is overlapping between any two nodes, the Oseen tensor formula should be modified as:[144]

$$T^0_{(i\alpha)(j\beta)} = (6\pi\eta\overline{\alpha})^{-1}\left[(1-\frac{9|r_{i,j}|}{32\overline{\alpha}}\sigma_{\alpha\beta}+\frac{3(r_{i,j}r_{i,j})_{\alpha\beta}}{32\overline{\alpha}|r_{i,j}|}\right] \tag{5.11}$$

where $\overline{\alpha}$ is the mean value of α_i and α_j.

The direct complete solution of Eq. 5.5 requires the diagonalization of a 6N×6N non-symmetric matrix, which is very complicated in the computational calculation. An alterative approach, so called perturbation analysis, is generally used to solve the Langevin equation and has proven to obtain results consistent with the complete langevin mode analysis.[144,149,150] The perturbation solution of the Langevin equation is performed as follows: The friction-free normal mode analysis yields the eigenvalues ω^2 and eigenvectors U. The friction matrix obtained from Oseen tensor conversion is then decomposed into diagonal γ and off-diagonal ε parts, the latter is treated as a perturbation. The "damping constant" corresponding to i^{th} normal mode is defined as $\eta_{i\alpha}=(U^T\gamma U)_{(i\alpha)(i\alpha)}$. The zero-order eigenvalues of the Langevin mode problem are obtained by ignoring the perturbation ε in friction matrix:[144]

$$\lambda_0 = \frac{1}{2}\begin{pmatrix}-\eta+\Delta & 0\\ 0 & -\eta-\Delta\end{pmatrix} \tag{5.12}$$

where Δ is a 3N diagonal matrix with element $\Delta_{i\alpha}=(\eta_{i\alpha}^2-4\omega_{i\alpha}^2)^{1/2}$. The function Δ determines the damping degree of each independent normal mode: if $\Delta_{i\alpha}$ is real, then the motion of $(i\alpha)^{th}$ mode is overdamped by water molecule. In other words, the motion is

restrained gradually to zero amplitude. Otherwise the mode is underdamped in which case the moving subject reaches to its equilibrium position or state quickly but then oscillates around it.

The time-average correlation function in Langevin mode is modified on the basis of Eq.5.3:

$$\left\langle \Delta r_{i\alpha}(t) \bullet \Delta r_{j\alpha}(0) \right\rangle = \frac{k_B T}{\gamma} \sum_{n=1}^{3N-6} \frac{U_{(i\alpha)n} U_{(j\alpha)n}}{\omega_n^{\,2}} \theta_n(t) \tag{5.13}$$

where $\theta(t)$ is a function:

$$\theta_i(t) = e^{-\eta_i t/2} \left[\cosh(\frac{\Delta_i t}{2}) + \frac{\eta_i}{\Delta_i} \sinh(\frac{\Delta_i t}{2}) \right] \tag{5.14}$$

5.3 Results

5.3.1 Homodimeric kinesin

We analyzed the dynamics of the dimeric as well as monomeric kinesins by the Langevin mode analysis coupled with ANM (dimeric kinesin: PDB code 3KIN,[104] monomeric kinesins: PDB code 2KIN[127] and 1MKJ[93]). The crystal structure of 3KIN consists of 708 residues whereas 2KIN structure contains 338 residues and 1MKJ 336 residues. In ANM, we set the force constant as 1.0 kcal/(mol·Å^2), as suggested by the previous studies in numerous proteins.[50] Then we calculated the mean-square fluctuations (self-correlation functions) of C_α atoms and then compared to the experimental B-factors to determine the value of cutoff distance r_c for all three systems under study. The results are shown in Fig. 5.1. One can observe from the figure that for the dimeric kinesin (3KIN, Fig. 5.1 (a)), the B-factor curve calculated with the r_c value

13.5 Å is closest to the experimental data. By using the same method, the r_c is set to 11 Å for the monomeric kinesins (both 2KIN and 1MKJ systems).

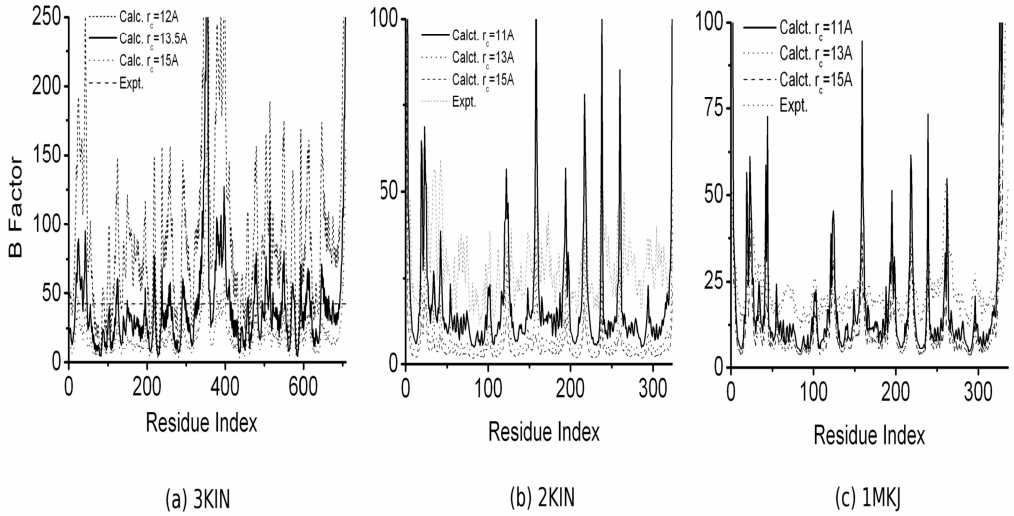


Figure 5.1. Cutoff distance dependence of the B-factor calculated for (a) dimeric kinesin (3KIN) and monomeric kinesins of (b) 2KIN and (c) 1MKJ.

In Langevin mode analysis, the friction constant matrix is calculated by Eq. 5.8-5.11, where the accessible surface area (ASA) of each residue is obtained from website program GETAREA 1.1.[172] The combination of the friction constant and the corresponding normal mode yields the damping constant η ($\eta_i=(U\gamma U^{-1})_{ii}$). Figure 5.2 shows the damping constant as a function of the normal mode frequency for the dimeric kinesin in water solution at room temperature (the viscosity is 0.89cP). The straight line in Fig. 5.2 denotes the function $\gamma=2\omega$. Therefore the modes with the corresponding damping constant above the straight line are overdamped, otherwise the modes are

underdamped. From this figure we can see that the first three nonzero lowest-frequency modes as well as the 11^{th} one are overdamped and the remaining modes are underdamped.

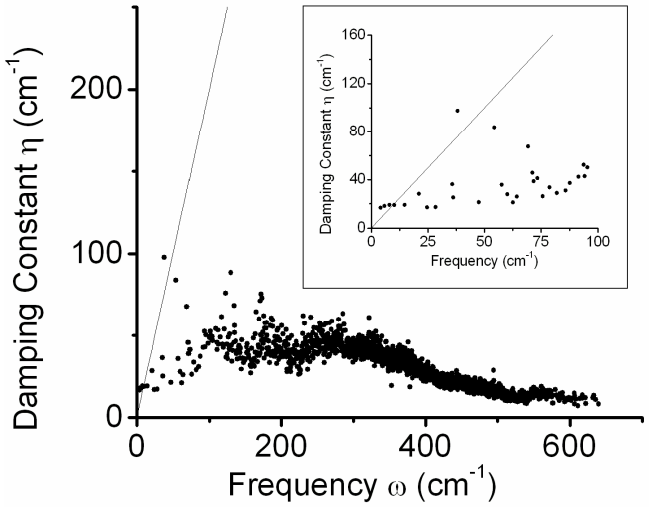


Figure 5.2. The damping constant as a function of the normal mode frequency for the dimeric kinesin in the water solution at room temperature. The straight line denotes the function $\gamma=2\omega$.

Figure 5.3 shows the real and imaginary parts of the Langevin mode zero-order eigenvalues λ_0 calculated from Eq. 5.12 for the dimeric kinesin.

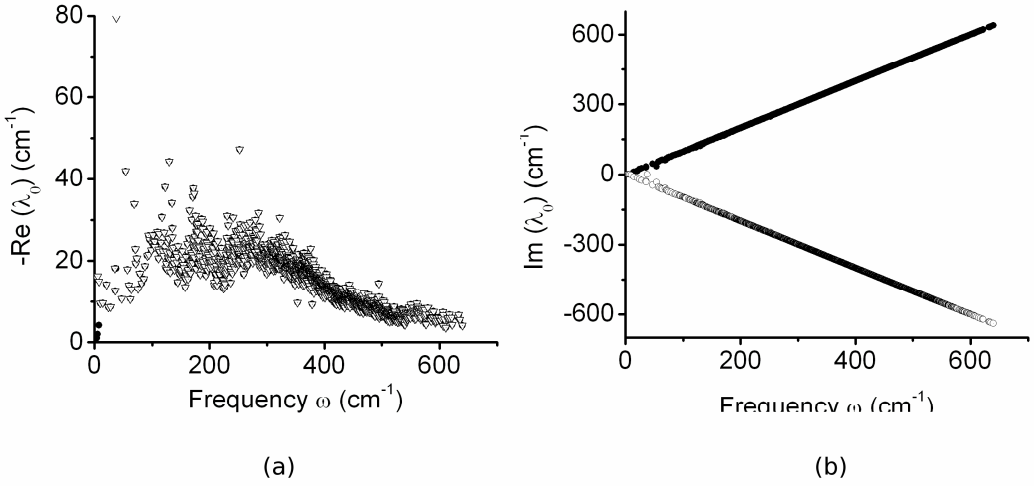


Figure 5.3. (a) The real part of the Langevin mode zero-order eigenvalues λ_0. (b) The imaginary part of λ_0 for the dimeric kinesin.

In the present study, only non-zero low-frequency modes are concerned, with the frequency range of 4~60 cm^{-1} (see Fig. 5.2 and Fig. 5.3). For dimeric kinesin, its six non-zero lowest-frequency normal modes exhibit distinct styles of kinesin motions whereas the motions from the modes with higher frequencies (7th-14th) are more or less the repetition of those in the former six modes, although in more disordered manners. The modes with even higher frequencies display disordered local fluctuations of C_α atoms. Figure 5.4 demonstrates snapshots for the motions of dimeric kinesin under different normal modes.

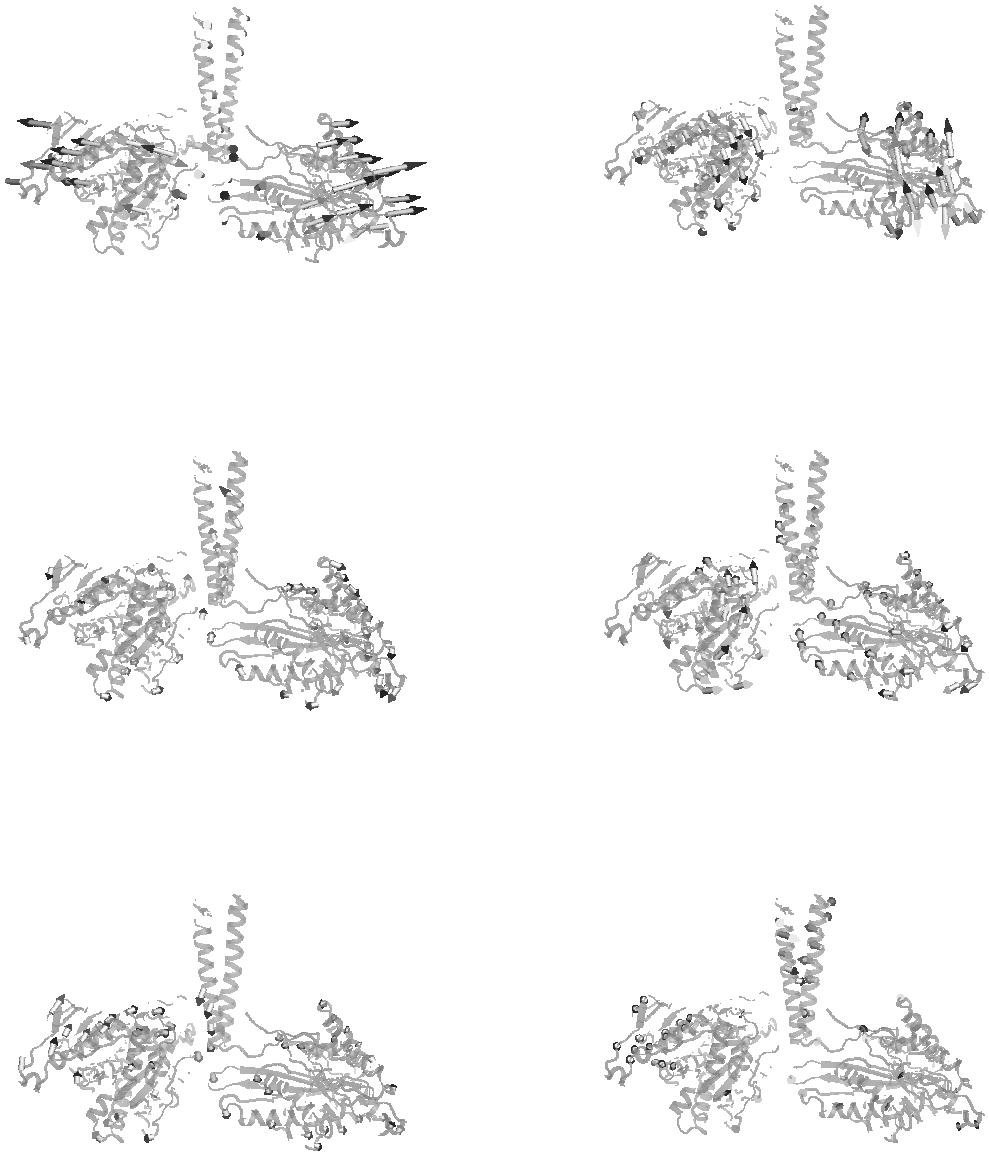


Figure 5.4. Snapshots of kinesin motions derived from the six non-zero lowest-frequency normal modes from ANM for the dimeric kinesin. Arrows indicate the motion directions of subdomains.

The detailed motions of the first six nonzero-frequency normal modes are described as follows: the first one (top left in Fig. 5.4) represents a symmetric bending motion of the two kinesin heads along the stalk. This mirror-symmetric motion is overdamped.

The second nonzero-frequency mode (top right in Fig. 5.4) represents a twist of the two heads in a symmetric manner: each head rotates through the juncture of the head and stalk. This motion occurs on the cross section of the kinesin heads, without any net movement of the heads in the forward/backward direction. The top of the stalk bends in the forward direction slightly. This point-symmetric motion of the two heads is also overdamped.

The third one (middle left in Fig. 5.4) is a bending motion of the two heads along z-direction (defined as the parallel direction of the stalk). This motion is most likely the flapping of the bird wings. As a result, the subdomains of the heads away from the stalk move downwards whereas the juncture of the heads and stalk rises up slightly. The two heads share same frequencies. Meanwhile the stalk keeps shaking with large scale between the heads. This motion again is overdamped.

In the fourth mode (middle right in Fig. 5.4), the two kinesin heads work asymmetrically. The left head plays a dominant role, which rolls through the juncture of this head and stalk. As a result, the portion of the head close to nucleotide-binding site moves forward. The right head moves in a similar manner but with a much smaller magnitude. Therefore this motion could be regarded as a kinesin “walking”: one head moves forward while the other one oscillates slightly around its original position, two heads play different roles. This ‘walking’ motion is underdamped.

The motion of kinesin heads under fifth normal mode (bottom left in Fig. 5.4) has the same fashion as that in third mode. However, the motion frequencies of two heads appear different. The sixth nonzero normal mode (bottom right in Fig. 5.4) represents the winding motion of the lower end of the coiled-coil stalk during which the kinesin heads keep relatively stable in structure. These two motions are underdamped.

We also checked the modes with larger frequencies (figures are not shown here). The motions could be distributed into the six categories as described above but in more complicated manners. The two heads work asymmetrically and individual residues behave disordered. All modes except 11^{th} one are underdamped. The latter mode represents a bending motion of the two kinesin heads, the same as that described by first mode, although in a much smaller magnitude.

5.3.2 Monomeric kinesins

The damping constant as a function of normal mode frequency was also calculated for monmeric kinesins and the results are shown in Fig. 5.5. The first three nonzero normal modes are overdamped for 1MKJ system (see Fig.5.5 (a)).

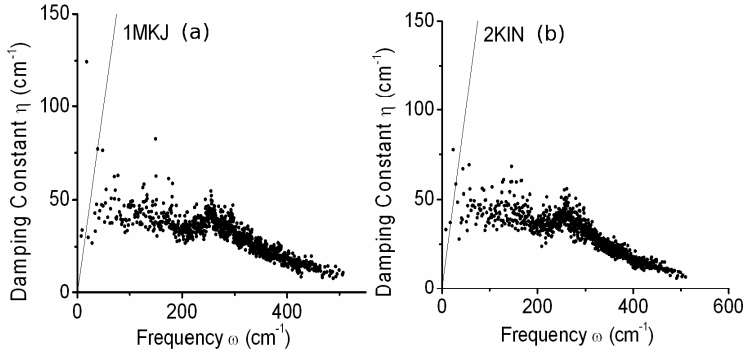


Figure 5.5. The damping constant as a function of the normal mode frequency for monomeric kinesins of (a) 1MKJ and (b) 2KIN in the water solution at room temperature. The straight line denotes the function $\gamma=2\omega$.

Through the careful analysis of normal modes, the first mode for 1MKJ refers to a "hook" motion: the single head bents in z-direction along the stalk. In the second mode, the head works similarly as the "walking" head in the fourth mode in dimeric kinesin: the head swings by itself in the forward direction. This can be treated as the "walking" of monomeric kinesin head. Two motions are represented in Fig. 5.6 (a & b). The third mode represents a motion similar as the "walking" motion, however with a much small magnitude. The global motions of the single head derived by the subsequent modes are rather disordered, especially in the part of the stalk.

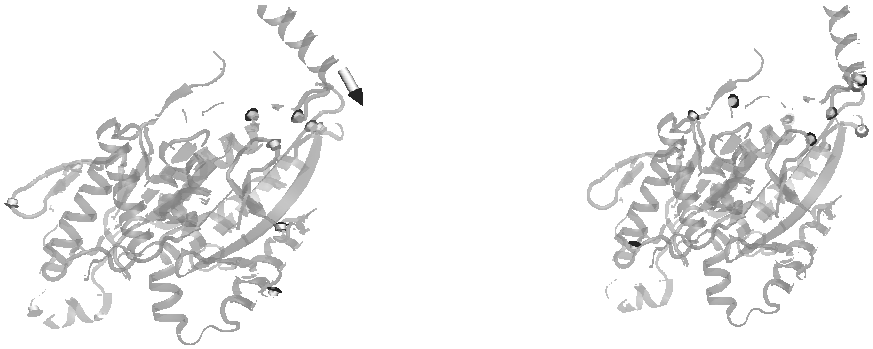


Figure 5.6. Two typical lowest-frequency normal model motions of the monomeric kinesin from the ANM calculation. Arrows indicate motion directions of subdomains.

As a comparison, the motion derived by the first nonzero normal mode of 2KIN is the "walking" motion of the single head as described above for 1MKJ whereas the second normal mode represents the "hook" motion. The third to fifth normal modes repeat the same "walking" motion but in more disordered manners. All normal modes except the first four slowest ones are underdamped.

5.3.3 Correlation functions

We indicated the mechanical coupling between the nucleotide binding site and the neck-linker in terms of the time-average cross-correlation function in Eq. 5.13. We calculated the cross-correlation functions between the residue Glu91 (P-loop) which connects to the nucleotide molecule (e.g., ADP) and each residue in the range of neck-linker (residues Arg323-Leu335) for both dimeric and monomeric kinesins, as shown in Fig. 5.7.

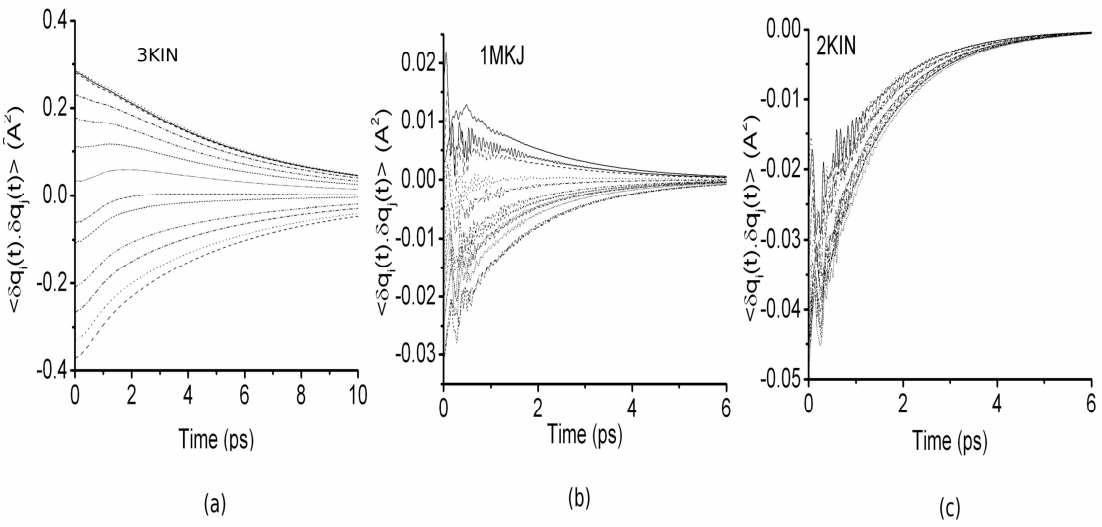


Figure 5.7. The cross-correlation function between the residue Glu91 in the nucleotide binding site and residues belonging to the neck-linker for (a) 3KIN and (b) 1MKJ and (c) 2KIN.

It is seen clearly that the absolute values of correlation functions between the P-loop residue and residues within neck-linker in dimeric kinesin are much larger compared to that in monomeric kinesins. Therefore, the correlation between the nucleotide binding site and the neck liner is strong in dimeric kinesin rather than in monomeric one. This observation supports the model that the power stroke of the neck-linkers induced by ATP binding is the fundamental driving force for the motility of kinesin dimer. On the contrary, the weak coupling between the nucleotide binding site and neck-linker in monomeric kinesin partially explains the reason why it can't walk on the microtubule processively.

5.4 Conclusion

To date, Langevin mode has been applied in several macromolecule systems[149,150]. Vibrational frequencies of normal modes are normally calculated from the principal component analysis (PCA) of molecular dynamics (MD) trajectories, e.g., using the program CHARMM or AMBER.[149,170,173] The calculation with all atoms involved limits the application of Langevin mode in large macromolecules. In the present study, ANM is used instead of standard normal mode. The combination of the Langevin mode analysis and ANM provides a plausible approach to describe the real large-scale dynamics of macromolecules in solvent. ANM not only produces eigenfrequency data compared well with those from the calculations using more accurate potential functions but also describes characteristic conformational changes of proteins by low-frequency normal modes. Through the comparison of the vibrational frequencies calculated in ANM with respect to the friction constants from Langevin mode analysis we could determine which normal mode is overdamped or underdamped. The dynamics study of proteins will not be limited in the vacuum condition.

In the present study we applied this method in kinesin system and obtain good results which are consistent with some novel experiments. In ANM, the parameters of cutoff distance r_c and force constant γ directly determine the magnitude of fluctuations. In order to obtain the precise eigenfrequencies comparable to the real vibrational motions, we set γ as 1.0 kcal/(mol$\cdot$Å^2) and determine the optimal value of r_c by fitting the calculated B factor curve to the experimental data for both dimeric and monomeric kinesins. For the dimeric kinesin, the motion derived from the forth slowest mode in

ANM is found to be a perfect simulation of the real movement of kinesin dimer on microtubule. This motion is underdamped whereas those modes with even lower frequency are overdamped. In those slower modes, the two kinesin heads work in symmetric manners: they bend in the mirror symmetry in first mode; they twist in the point symmetry in second mode and they flap in the mirror symmetry in third mode. These motions are against the functional asymmetry of the kinesin heads in the real circumstance and therefore the rate processes of them are restricted. As a comparison, in kinesin monomer systems, the slowest (3KIN) or second slowest (1MKJ) normal mode motion well mimics the 'walking' of the single head but is overdamped.

We observed the strong coupling between the nucleotide-binding site and neck-linker in dimeric kinesin rather than monomeric one. Cross-correlation functions between the nucleotide binding site and residues within the neck-linker in dimeric kinesin are obviously larger in comparison to those in monomeric kinesin. This might explain the experimental observation that the neck-linker trends to tilt in the forward direction once the ATP binds to its affiliated kinesin head and therefore dimeric kinesin walks on the microtubule processively; the monomeric kinesin, on the other hand, can never walks processively. This method is anticipated to be applied in more protein systems.

CHAPTER VI

SUMMARY

Molecular dynamics (MD) simulation has been widely used in the field of computational biology, including the studies on the protein folding and protein functions at atomic level. However the large computational cost of the conventional MD simulation severely limits its application to large biological systems. Strategies to reduce the computational cost include both the development of efficient sampling algorithms and the usage of implicit solvent models instead of explicit ones. Moreover, in order to investigate even larger systems, coarse-grained models have also been developed and applied in protein studies. All three strategies were employed in this dissertation to study the folding and functions of various proteins.

In the investigation of protein folding, we studied the folding mechanism of β-hairpins (peptide 1, GB1, TRPZIP2 and TRPZIP4) and β-sheets (20mer and 20mer^{D}P6D) using the integrated tempering sampling (ITS) method. We observed the significant differences among folding free energy landscapes of four hairpins which indicate the different folding mechanisms of similar β-structures of varied sequences. The folding mechanism of a β-structure is determined by its turn sequence and sidechain hydrophobicity. In the presence of strong β-turn promoting sequence, the β-hairpin folding is most likely to follow the modified "hydrogen-bond-centric" mechanism whereas the folding scenario is consistent with the "hydrophobic-core-centric" mechanism. It is also observed that the turn sequence and sidechain hydrophobicity affect the stability of their nearby backbone hydrogen bonds. The easy formation of the

turn structure (e.g., in peptide 1 and TRPZIP2) leads to the easy formation of the hydrogen bond closest to the turn, although it is broken quickly after the hydrophobic core formation. In contrast, the hydrogen bond closest to the turn structure exhibits the extreme instability through the assembling of backbone hydrogen bonds. On the other hand, strong hydrophobic interactions induce the stronger stability of the two hydrogen bonds in the middle of the hydrophobic core compared to other ones whereas the stability of hydrogen bonds except for the two marginal ones is even in the presence of weak hydrophobic interactions.

Two implicit solvent models (GB^{OBC} and GBn) were tested in our folding simulation of GB1 peptide. Through the quantitive thermodynamics analysis, we observed that the results obtained from GB^{OBC} model are consistent with those from the explicit solvent model (AMBER ff96/SPC). Therefore, in combination with AMBER ff96 force field, GB^{OBC} is better in describing the folding of β-hairpin compared to GBn.

In the investigation of protein functions, we studied the processive motion of kinesin at the mesoscopic level, which is a family member of protein motors. A simple model was proposed to explain the asymmetric hand-over-hand mechanism of the kinesin walking on the microtubule. The asymmetry property of the hand-over-hand motion of kinesin is proposed in this model to be induced by the length change of neck-linkers during the kinesin walking. The external force and ATP concentration dependence of the walking speed, the forward/backward step ratio and the dwell times for both forward and backward steps were calculated for both conventional kinesin and kinesin mutants and the results were compared to experiments. Meanwhile the global conformational change

of each kinesin head which occurs in each step of kinesin walking was studied by using the anisotropic network model (ANM).

REFERENCES

(1) Lodish, H.; Berk, A.; Matsudaira, P.; Kaiser, C. A.; Krieger, M.; Scott, M. P.; Zipurksy, S. L.; Darnell, J. *Molecular Cell Biology*; 5th ed.; WH Freeman and Company: New York, NY, **2004**.

(2) Ridley, M. *Genome: The Autobiography of a Specials in 23 Chapters*; Harper Perennial: New York, NY, **2006**.

(3) Dobson, C. M. *Nature* **2003**, *426*, 884-890.

(4) Morra, G.; Meli, M.; Colombo, G. *Curr. Protein Pept. Sci.* **2008**, *9*, 181-196.

(5) Du, D. G.; Zhu, Y. J.; Huang, C. Y.; Gai, F. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 15915-15920.

(6) Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 5578-5583.

(7) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. *Nature* **1958**, *181*, 662-666.

(8) Scapin, G. *Curr. Pharm. Design* **2006**, *12*, 2087-2097.

(9) Keiderling, T. A. *Curr. Opin. Chem. Biol.* **2002**, *6*, 682-688.

(10) Karplus, M.; Kuriyan, J. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6679-6685.

(11) Scheraga, H. A.; Khalili, M.; Liwo, A. *Annu. Rev. Phys. Chem.* **2007**, *58*, 57-83.

(12) A.R.Leach *Molecular Modeling Principles and Applications*; 2nd ed.; Pearson Education Ltd.: Harlow, England, **2001**.

(13) Hockney, R. W. *Methods in Computational Physics*; Academic Press: New York, NY, **1970**.

(14) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. *J. Chem. Phys.* **1982**, *76*, 637-649.

(15) Beeman, D. *J. Comput. Phys.* **1976**, *20*, 130-139.

(16) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187-217.

(17) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1-41.

(18) Berendsen, H. J. C.; Vanderspoel, D.; Vandrunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43-56.

(19) Higo, J. *Lect. notes comp. sci.* **2008**, *4759,* 374-382.

(20) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005-3014.

(21) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am Chem. Soc.* **1990**, *112*, 6127-6129.

(22) Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. *J. Chem. Theory Comput.* **2007**, *3*, 156-169.

(23) Gilson, M. K.; Davis, M. E.; Luty, B. A.; Mccammon, J. A. *J. Phys. Chem.* **1993**, *97*, 3591-3600.

(24) Onufriev, A.; Bashford, D.; Case, D. A. *J. Phys. Chem. B* **2000**, *104*, 3712-3720.

(25) Miller, C. A.; Gellman, S. H.; Abbott, N. L.; de Pablo, J. J. *Biophys. J.* **2008**, *95*, 3123-3136.

(26) Roe, D. R.; Okur, A.; Wickstrom, L.; Hornak, V.; Simmerling, C. *J. Phys. Chem. B* **2007**, *111*, 1846-1857.

(27) Shao, Q.; Yang, L. J.; Gao, Y. Q. *J. Chem. Phys.* **2009**, *130*, 195104/1-195104/6.

(28) Shell, M. S.; Ritterson, R.; Dill, K. A. *J. Phys. Chem. B* **2008**, *112*, 6878-6886.

(29) Zhou, R. H. *Proteins* **2003**, *53*, 148-161.

(30) Frantz, D. D.; Freeman, D. L.; Doll, J. D. *J. Chem. Phys.* **1990**, *93*, 2769-2784.

(31) Berg, B. A.; Neuhaus, T. *Phys. Lett. B* **1991**, *267*, 249-253.

(32) Piela, L.; Kostrowicki, J.; Scheraga, H. A. *J. Phys. Chem.* **1989**, *93*, 3339-3346.

(33) Grubmuller, H. *Phys. Rev. E* **1995**, *52*, 2893-2906.

(34) Voter, A. F. *Phys. Rev. Lett.* **1997**, *78*, 3908-3911.

(35) Bartels, C.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 865-880.

(36) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141-151.

(37) Sugita, Y.; Kitao, A.; Okamoto, Y. *J. Chem. Phys.* **2000**, *113*, 6042-6051.

(38) Lee, J.; Scheraga, H. A.; Rackovsky, S. *J. Comput. Chem.* **1997**, *18*, 1222-1232.

(39) Itoh, S. G.; Okamoto, Y. *J. Chem. Phys.* **2006**, *124*, 104103/1-104103/14.

(40) Bussi, G.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2006**, *96*, 090601-090604.

(41) Berne, B. J.; Straub, J. E. *Curr. Opin. Struct. Biol.* **1997**, *7*, 181-189.

(42) Gao, Y. Q.; Yang, L. J. *J. Chem. Phys.* **2006**, *125*, 114103/1-114103/5.

(43) Gao, Y. Q. *J. Chem. Phys.* **2008**, *128*, 064105/1-064105/5.

(44) Levitt, M.; Warshel, A. *Nature* **1975**, *253*, 694-698.

(45) Kolinski, A.; Ilkowski, B.; Skolnick, J. *Biophys. J.* **1999**, *77*, 2942-2952.

(46) Chan, H. S.; Dill, K. A. *Proteins* **1996**, *24*, 335-344.

(47) Dill, K. A.; Chan, H. S. *Nat. Struct. Biol.* **1997**, *4*, 10-19.

(48) Liwo, A.; Khalili, M.; Scheraga, H. A. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 2362-2367.

(49) Tirion, M. M. *Phys. Rev. Lett.* **1996**, *77*, 1905-1908.

(50) Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. *Biophys. J.* **2001**, *80*, 505-515.

(51) Kondrashov, D. A.; Cui, Q.; Phillips, G. N. *Biophys. J.* **2006**, *91*, 2760-2767.

(52) Alonso, D. O. V.; Daggett, V. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 133-138.

(53) Blanco, F. J.; Rivas, G.; Serrano, L. *Nat. Struct. Biol.* **1994**, *1*, 584-590.

(54) Duan, Y.; Kollman, P. A. *Science* **1998**, *282*, 740-744.

(55) Galzitskaya, O. V.; Higo, J.; Finkelstein, A. V. *Curr. Protein Pept. Sci.* **2002**, *3*, 191-200.

(56) Munoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196-199.

(57) Kubelka, J.; Hofrichter, J.; Eaton, W. A. *Curr. Opin. Struct. Biol.* **2004**, *14*, 76-88.

(58) Zerella, R.; Evans, P. A.; Ionides, J. M. C.; Packman, L. C.; Trotter, B. W.; Mackay, J. P.; Williams, D. H. *Protein Sci.* **1999**, *8*, 1320-1331.

(59) Silva, R. A. G. D.; Sherman, S. A.; Keiderling, T. A. *Biopolymers* **1999**, *50*, 413-423.

(60) Bonomi, M.; Branduardi, D.; Gervasio, F. L.; Parrinello, M. *J. Am. Chem. Soc.* **2008**, *130*, 13938-13944.

(61) Klimov, D. K.; Thirumalai, D. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 2544-2549.

(62) Pande, V. S.; Rokhsar, D. S. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9062-9067.

(63) Roccatano, D.; Amadei, A.; Di Nola, A.; Berendsen, H. J. C. *Protein Sci.* **1999**, *8*, 2130-2143.

(64) Zhou, R. H.; Berne, B. J.; Germain, R. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 14931-14936.

(65) Dinner, A. R.; Lazaridis, T.; Karplus, M. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9068-9073.

(66) Tsai, J.; Levitt, M. *Biophys. Chem.* **2002**, *101*, 187-201.

(67) Sanbonmatsu, K. Y.; Garcia, A. E. *Biophys. J.* **2001**, *80*, 399a-399a.

(68) Garcia, A. E.; Sanbonmatsu, K. Y. *Proteins* **2001**, *42*, 345-354.

(69) Nguyen, P. H.; Stock, G.; Mittag, E.; Hu, C. K.; Li, M. S. *Proteins* **2005**, *61*, 795-808.

(70) Wei, G. H.; Mousseau, N.; Derreumaux, P. *Proteins* **2004**, *56*, 464-474.

(71) Yoda, T.; Sugita, Y.; Okamoto, Y. *Proteins* **2007**, *66*, 846-859.

(72) Espinosa, J. F.; Munoz, V.; Gellman, S. H. *J. Mol. Biol.* **2001**, *306*, 397-402.

(73) Munoz, V.; Henry, E. R.; Hofrichter, J.; Eaton, W. A. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 5872-5879.

(74) Santiveri, C. M.; Rico, M.; Jimenez, M. A. *Protein Sci.* **2000**, *9*, 2151-2160.

(75) Wang, H. W.; Sung, S. S. *Biopolymers* **1999**, *50*, 763-776.

(76) Carr, J. M.; Wales, D. J. *J. Phys. Chem. B* **2008**, *112*, 8760-8769.

(77) Ferrara, P.; Caflisch, A. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 10780-10785.

(78) Searle, M. S.; Ciani, B. *Curr. Opin. Struct. Biol.* **2004**, *14*, 458-464.

(79) Syud, F. A.; Stanger, H. E.; Mortell, H. S.; Espinosa, J. F.; Fisk, J. D.; Fry, C. G.; Gellman, S. H. *J. Mol. Biol.* **2003**, *326*, 553-568.

(80) Hughes, R. M.; Waters, M. L. *Curr. Opin. Struct. Biol.* **2006**, *16*, 514-524.

(81) Rotondi, K. S.; Gierasch, L. M. *Biopolymers* **2006**, *84*, 13-22.

(82) Roe, D. R.; Hornak, V.; Simmerling, C. *J. Mol. Biol.* **2005**, *352*, 370-381.

(83) Trevino, S. R.; Schaefer, S.; Scholtz, J. M.; Pace, C. N. *J. Mol. Biol.* **2007**, *373*, 211-218.

(84) Schenck, H. L.; Gellman, S. H. *J. Am. Chem. Soc.* **1998**, *120*, 4869-4870.

(85) Chen, P. Y.; Lin, C. K.; Lee, C. T.; Jan, H.; Chan, S. I. *Protein Sci.* **2001**, *10*, 1794-1800.

(86) Kuo, N. N. W.; Huang, J. J. T.; Miksovska, J.; Chen, R. P. Y.; Larsen, R. W.; Chan, S. I. *J. Am. Chem. Soc.* **2005**, *127*, 16945-16954.

(87) Kuznetsov, S. V.; Hilario, J.; Keiderling, T. A.; Ansari, A. *Biochemistry* **2003**, *42*, 4321-4332.

(88) Hirokawa, N. *Science* **1998**, *279*, 519-526.

(89) Vale, R. D. *Cell* **2003**, *112*, 467-480.

(90) Cross, R. A. *Trends Biochem. Sci.* **2004**, *29*, 301-309.

(91) Asbury, C. L. *Curr. Opin. Cell Biol.* **2005**, *17*, 89-97.

(92) Yildiz, A.; Selvin, P. R. *Trends Cell Biol.* **2005**, *15*, 112-120.

(93) Sindelar, C. V.; Budny, M. J.; Rice, S.; Naber, N.; Fletterick, R.; Cooke, R. *Nat. Struct. Biol.* **2002**, *9*, 844-848.

(94) Rice, S.; Lin, A. W.; Safer, D.; Hart, C. L.; Naber, N.; Carragher, B. O.; Cain, S. M.; Pechatnikova, E.; Wilson-Kubalek, E. M.; Whittaker, M.; Pate, E.; Cooke, R.; Taylor, E. W.; Milligan, R. A.; Vale, R. D. *Nature* **1999**, *402*, 778-784.

(95) Tomishige, M.; Stuurman, N.; Vale, R. *Nat. Struct. Mol. Biol.* **2006**, *13*, 887-894.

(96) Kull, F. J.; Endow, S. A. *J. Cell Sci.* **2002**, *115*, 15-23.

(97) Rosenfeld, S. S.; Xing, J.; Jefferson, G. M.; Cheung, H. C.; King, P. H. *J. Biol. Chem.* **2002**, *277*, 36731-36739.

(98) Hirose, K.; Lockhart, A.; Cross, R. A.; Amos, L. A. *Nature* **1995**, *376*, 277-279.

(99) Yun, M.; Bronner, C. E.; Park, C. G.; Cha, S. S.; Park, H. W.; Endow, S. A. *EMBO J.* **2003**, *22*, 5382-5389.

(100) Asenjo, A. B.; Krohn, N.; Sosa, H. *Nat. Struct. Biol.* **2003**, *10*, 836-842.

(101) Kawaguchi, K.; Ishiwata, S. *Science* **2001**, *291*, 667-669.

(102) Kawaguchi, K.; Uemura, S.; Ishiwata, S. *Biophys. J.* **2003**, *84*, 1103-1113.

(103) Uemura, S.; Kawaguchi, K.; Yajima, J.; Edamatsu, M.; Toyoshima, Y. Y.; Ishiwata, S. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 5977-5981.

(104) Kozielski, F.; Sack, S.; Marx, A.; Thormahlen, M.; Schonbrunn, E.; Biou, V.; Thompson, A.; Mandelkow, E. M.; Mandelkow, E. *Cell* **1997**, *91*, 985-994.

(105) Hackney, D. D. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 6314-6318.

(106) Ma, Y. Z.; Taylor, E. W. *J. Biol. Chem.* **1997**, *272*, 717-723.

(107) Coy, D. L.; Wagenbach, M.; Howard, J. *J. Biol. Chem.* **1999**, *274*, 3667-3671.

(108) Nishiyama, M.; Muto, E.; Inoue, Y.; Yanagida, T.; Higuchi, H. *Nat. Cell Biol.* **2001**, *3*, 425-428.

(109) Schnitzer, M. J.; Block, S. M. *Nature* **1997**, *388*, 386-390.

(110) Schnitzer, M. J.; Visscher, K.; Block, S. M. *Nat. Cell. Biol.* **2000**, *2*, 718-723.

(111) Visscher, K.; Schnitzer, M. J.; Block, S. M. *Nature* **1999**, *400*, 184-189.

(112) Yildiz, A.; Tomishige, M.; Vale, R. D.; Selvin, P. R. *Science* **2004**, *303*, 676-678.

(113) Block, S. M.; Svoboda, K. *Biophys. J.* **1995**, *68*, S230-S241.

(114) Hua, W.; Chung, J.; Gelles, J. *Science* **2002**, *295*, 844-848.

(115) Nishiyama, M.; Higuchi, H.; Yanagida, T. *Nat. Cell. Biol.* **2002**, *4*, 790-797.

(116) Carter, N. J.; Cross, R. A. *Nature* **2005**, *435*, 308-312.

(117) Coppin, C. M.; Finer, J. T.; Spudich, J. A.; Vale, R. D. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 1913-1917.

(118) Asbury, C. L.; Fehr, A. N.; Block, S. M. *Science* **2003**, *302*, 2130-2134.

(119) Higuchi, H.; Bronner, C. E.; Park, H. W.; Endow, S. A. *EMBO J.* **2004**, *23*, 2993-2999.

(120) Thoresen, T.; Gelles, J. *Biophys. J.* **2004**, *86*, 409-409.

(121) Inoue, Y.; Iwane, A. H.; Miyai, T.; Muto, E.; Yanagida, T. *Biophys. J.* **2001**, *81*, 2838-2850.

(122) Fisher, M. E.; Kolomeisky, A. B. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 7748-7753.

(123) Kolomeisky, A. B.; Stukalin, E. B.; Popov, A. A. *Phys. Rev. E* **2005**, *71*, 031902/1-031902/8.

(124) Astumian, R. D. *Philos. Trans. R. Soc. London, Ser. B* **2000**, *355*, 511-522.

(125) Peskin, C. S.; Oster, G. *Biophys. J.* **1995**, *68*, S202-S211.

(126) Thomas, N.; Imafuku, Y.; Kamiya, T.; Tawada, K. *Philos. Trans. R. Soc. London, Ser. B* **2002**, *269*, 2363-2371.

(127) Sack, S.; Muller, J.; Marx, A.; Thormahlen, M.; Mandelkow, E. M.; Brady, S. T.; Mandelkow, E. *Biochemistry* **1997**, *36*, 16155-16165.

(128) Levitt, M.; Sander, C.; Stern, P. S. *J. Mol. Biol.* **1985**, *181*, 423-447.

(129) Brooks, B.; Karplus, M. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 6571-6575.

(130) Tirion, M. M.; Avraham, D. B.; Holmes, K. *Biophys. J.* **1993**, *64*, A146-A146.

(131) Tirion, M. M.; Benavraham, D. *J. Mol. Biol.* **1993**, *230*, 186-195.

(132) Chennubhotla, C.; Rader, A. J.; Yang, L. W.; Bahar, I. *Phys. Biol.* **2005**, *2*, S173-S180.

(133) Eyal, E.; Yang, L. W.; Bahar, I. *Bioinformatics* **2006**, *22*, 2619-2627.

(134) Ming, D.; Kong, Y. F.; Lambert, M. A.; Huang, Z.; Ma, J. P. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 8620-8625.

(135) He, J. B.; Zhang, Z. Y.; Shi, Y. Y.; Liu, H. Y. *J. Chem. Phys.* **2003**, *119*, 4005-4017.

(136) Zheng, W. J.; Brooks, B. R. *Biophys. J.* **2005**, *89*, 167-178.

(137) Zheng, W. J.; Brooks, B. R. *Biophys. J.* **2005**, *88*, 3109-3117.

(138) Zheng, W. J.; Doniach, S. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 13253-13258.

(139) Tama, F. *Protein Peptide Lett.* **2003**, *10*, 119-132.

(140) Tama, F.; Sanejouand, Y. H. *Protein Eng.* **2001**, *14*, 1-6.

(141) Kundu, S.; Jernigan, R. L. *Biophys. J.* **2004**, *86*, 3846-3854.

(142) Isin, B.; Doruker, P.; Bahar, I. *Biophys. J.* **2002**, *82*, 569-581.

(143) Kitao, A.; Hirata, F.; Go, N. *Chem. Phys.* **1991**, *158*, 447-472.

(144) Lamm, G.; Szabo, A. *J. Chem. Phys.* **1986**, *85*, 7334-7348.

(145) Stilling.Fh; Rahman, A. *J. Chem. Phys.* **1974**, *60*, 1545-1557.

(146) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269-6271.

(147) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926-935.

(148) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. *J. Chem. Phys.* **2004**, *120*, 9665-9678.

(149) Ansari, A. **1999**, *J. Chem. Phys. 110*, 1774-1780.

(150) Kottalam, J.; Case, D. A. *Biopolymers* **1990**, *29*, 1409-1421.

(151) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824-19839.

(152) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383-394.

(153) Gronenborn, A. M.; Clore, G. M. *Science* **1991**, *254*, 581-582.

(154) Gao, Y. Q.; Yang, L. J.; Fan, Y. B.; Shao, Q. *Int. Rev. Phys. Chem.* **2008**, *27*, 201-227.

(155) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.

(156) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327-341.

(157) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684-3690.

(158) Zaman, M. H.; Shen, M. Y.; Berry, R. S.; Freed, K. F.; Sosnick, T. R. *J. Mol. Biol.* **2003**, *331*, 693-711.

(159) Gnanakaran, S.; Garcia, A. E. *Proteins* **2005**, *59*, 773-782.

(160) Ozkan, S. B.; Wu, G. A.; Chodera, J. D.; Dill, K. A. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 11987-11992.

(161) Yang, L. J.; Shao, Q.; Gao, Y. Q. *J. Phys. Chem. B* **2009**, *113*, 803-808.

(162) Yang, L. J.; Shao, Q.; Gao, Y. Q. *J. Chem. Phys.* **2009**, *130*, 124111/1-124111/8.

(163) Streicher, W. W.; Makhatadze, G. I. *J. Am. Chem. Soc.* **2006**, *128*, 30-31.

(164) Shao, Q.; Yang, L. J.; Gao, Y. Q. *J. Chem. Phys.* **2009**, *130*, 195104/1-195104/6.

(165) Kyte, J.; Doolittle, R. F. *J. Mol. Biol.* **1982**, *157*, 105-132.

(166) Shao, Q.; Gao, Y. Q. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8072-8077.

(167) Shao, Q.; Gao, Y. Q. *Biochemistry* **2007**, *46*, 9098-9106.

(168) Asenjo, A. B.; Weinberg, Y.; Sosa, H. *Nat. Struct. Mol. Biol.* **2006**, *13*, 648-654.

(169) Cross, R. A.; Crevel, I.; Carter, N. J.; Alonso, M. C.; Hirose, K.; Amos, L. A. *Philos. Trans. R. Soc. Landon, Ser. B* **2000**, *355*, 459-464.

(170) Moritsugu, K.; Smith, J. C. *J. Phys. Chem. B* **2005**, *109*, 12182-12194.

(171) Carrasco, B.; de la Torre, J. G.; Zipper, P. *Eur. Biophys. J. Biophys. Lett.* **1999**, *28*, 510-515.

(172) Fraczkiewicz, R.; Braun, W. *J. Comput. Chem.* **1998**, *19*, 319-333.

(173) Kitao, A.; Go, N. *Curr. Opin. Struct. Biol.* **1999**, *9*, 164-169.

VITA

Name: Qiang Shao

Address: Department of Chemistry, 3255 TAMU, College Station, TX, 77843

Email Address: qshao@mail.chem.tamu.edu

Education: B.S., Chemistry, Nanjing University, China, 1999
M.A., Chemistry, College of William & Mary, 2004
Ph.D., Chemistry, Texas A&M University, 2009

Publications:

1. Shao Q.; Wei H.; Gao Y. Q. *J. Amer. Chem. Soc.* **2009***,* submitted.

2. Gao Y. Q.; Shao Q. Book chapter in *Proteins Energy Heat & Signal Flow*; CRC Press: Boca Raton, FL, **2009**.

3. Fan Y. B.; Lund L. L.; Shao Q.; Gao Y. Q.; Raushel F. M. *J. Amer. Chem. Soc.* **2009**, *131*, 10211-10219.

4. Shao Q.; Yang, L. J.; Gao Y. Q. *J. Chem. Phys.* **2009**, *130*, 195104/1-195104/6.

5. Yang, L. J.; Shao Q.; Gao Y. Q. *J. Chem. Phys.* **2009**, *130*, 124111/1-124111/8.

6. Yang, L. J.; Shao Q.; Gao Y. Q. *J. Phys. Chem. B* **2009**, *113*, 803-808.

7. Gao Y. Q.; Yang L. J.; Fan Y. B.; Shao Q. *Int. Rev. Phys. Chem.* **2008**, *27*, 201-227.

8. Shao Q.; Gao Y.Q. *Biochemistry* **2007**, *46*, 9098-9106.

9. Shao Q.; Gao Y.Q. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8072-8077.